
W0934

(TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MParch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 31 11:03:21 2000; MasPar time 8.27 Seconds
Tabular output not generated. 638.314 Million cell updates/sec

Title: >US-09-186-342-2
Description: (1-223) from US09186342.pap
Perfect Score: 1686
Sequence: 1 MRVSGVRLRLALIFAIVTTW.....KYEGWPELLEMEGCMPPKPF 223

Scoring table: PAM 150
Gap 11

Searched: 188963 seqs, 23686106 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-geneseq36
1:geneseqp

Statistics: Mean 33.108; Variance 140.762; scale 0.235

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	ID	Description	Pred. No.
1	1686	100.0	223	Protein encoded by a c	1.51e-153
2	993	58.9	135	Protein sequence encod	2.59e-84
3	993	58.9	135	Human colon specific g	2.59e-84
4	108	6.4	618	DmORF1 potassium chann	2.45e+00
5	102	6.0	993	Human mitochondrial is	6.95e+00
6	95	5.6	187	Sequence of yellow tai	2.28e+01
7	95	5.6	476	Gentian flavonoid 3'-h	2.28e+01
8	94	5.6	642	Human nuclear protein	2.70e+01
9	94	5.6	642	Amino acid sequence of	2.70e+01
10	92	5.5	146	Haemin-binding protein	3.77e+01
11	92	5.5	176	Haemin-binding protein	3.77e+01
12	92	5.5	178	Haemin-binding protein	3.77e+01
13	91	5.4	191	Bovine Somatotropin an	4.44e+01
14	91	5.4	217	Bovine Somatotropin an	4.44e+01
15	91	5.4	217	Bovine growth hormone	4.44e+01
16	91	5.4	533	HUG-Br1.	4.44e+01
17	91	5.4	534	HUG-Br2.	4.44e+01
18	89	5.3	191	Bovine Somatotropin an	6.17e+01
19	89	5.3	204	Yellow tail/fin tuna g	6.17e+01
20	89	5.3	204	Yellowtail tuna growth	6.17e+01
21	90	5.3	217	Bovine growth hormone	5.24e+01
22	90	5.3	217	Bovine growth hormone	5.24e+01
23	89	5.3	275	Staphylococcus aureus	6.17e+01

RESULT 1

ID W80475 standard; Protein; 223 AA.
AC W80475;
DT 28-JAN-1999 (first entry)
DE Protein encoded by a consensus sequence of CS141 EST.
KW EST sequence; CS141; gastrointestinal tract; cancer; Barret's esophagus;
KW gastric ulcer; gastritis; leiomyoma; polypos; Crohn's disease;
KW ulcerative colitis; pancreatitis.
OS Synthetic.
OS Homo sapiens.
PN W09844133-AT.
PD 08-OCT-1998.
PF 31-MAR-1998; U06337.
PR 31-MAR-1997; US-828489.
PA (ABBO) ABBOTT LAB.
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
PI Gordon J, Grados EN, Hayden M, Hodges SC, Klass MR,
PI Kratochvil JB, Roberts-Rapp L, Russell JC, Stroupe SD;
DR WPI; 98-568280/48.
DR N-PSDB; V63516.
PT New gastrointestinal tract specific polynucleotides, CS141 - used to
develop products for the diagnosis and treatment of e.g. cancers,
gastric ulcer, gastritis, Crohn's disease, ulcerative colitis or
pancreatitis

Claim 17; Pages 95-96; 116pp; English.

FS The present sequence is encoded by a consensus sequence of a set of
contiguous and partially overlapping EST sequences designated CS141.
CC The sequences are isolated from a cDNA library made from gastrointestinal
tract tumour and normal tissues. The CS141 gene is useful as a marker for
gastrointestinal tract disorders. The methods and products can be
used in detecting, diagnosing, staging, monitoring, prognosticating,
preventing, or determining the predisposition to diseases and
conditions of the gastrointestinal tract, such as gastrointestinal
tract cancers, Barret's esophagus, gastric ulcer, gastritis,
leiomyoma, polypos, Crohn's disease, ulcerative colitis, pancreatitis.
SQ Sequence 223 AA;

Query Match 100.0%; Score 1686; DB 1; Length 223;

Best Local Similarity 100.0%; Pred. No. 1.51e-153;

Matches 223; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 1 MRVSGVRLRLALIFAIVTTWFIKSYMSFKMTIRLPWLASPTKEIQVKYKGLIKPC 60

Qy 1 MRVSGVRLRLALIFAIVTTWFIKSYMSFKMTIRLPWLASPTKEIQVKYKGLIKPC 60

Db 61 PANYFAFKICSAANVVGPTCFEDRMIMSPVKNVGRGLIALVNGTGTGAVLGOKAFDM 120

|||||

QY 61 PANYFAFKICSGAANVVTGTCEDRMIMSPVKNVGRGLNLTALVNGTTGAVLGOKAFDM 120
Db 121 YSGDVHMLVKFLKEIPGGALVLVASYDDPGTKMNDSEKRLKFLSDLGSSYAKQLGFRDSWVF 180
QY 121 YSGDVHMLVKFLKEIPGGALVLVASYDDPGTKMNDSEKRLKFLSDLGSSYAKQLGFRDSWVF 180
Db 181 IGAKDLRGKSPFEQFLKNSPDNTNKYEGWPELLEMEGCMPPKPF 223
QY 181 IGAKDLRGKSPFEQFLKNSPDNTNKYEGWPELLEMEGCMPPKPF 223

RESULT 2

ID W46878 standard; Protein; 135 AA.
AC W46878;
DT 22-JUN-1998 (first entry)
DE Protein sequence encoded by a colon-specific gene.
KW Colon-specific gene; probe: detection; expression; human;
KW diagnostic assay; colon cancer; antibody; screening.
OS Homo sapiens.
PN US5733748-A.
PD 31-MAR-1998.
PF 06-JUN-1995; 469667.
PR 06-JUN-1995; US-469667.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen C, Yu G;
DR WPI: 98-729823/20.
DR N-PSDB; V16671.
PT Colon-specific nucleic acids - useful as probes for detecting colon
PT cancer micrometastases
PS Claim 1; Fig 4; 51pp; English.
CC W46876-80 and W4682-85 represent proteins encoded by colon-specific
CC genes. The polynucleotides encoding these proteins can be used
CC as probes to detect expression of the corresponding human genes,
CC e.g. in diagnostic assays for detecting micrometastases of colon cancer.
CC Recombinant cells containing the polynucleotides can be used to
CC produce the proteins, in order that antibodies can be raised and
CC used in further screening or diagnostics.
SQ Sequence 135 AA;

Query Match 58.9%; Score 993; DB 1; Length 135;
Best Local Similarity 100.0%; Pred. No. 2.59e-84;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 MSPVKNVGRGLNLTALVNGTTGAVLGOKAFDMYSGDVHMLVKFLKEIPGGALVLVASYDD 60
QY 89 MSPVKNVGRGLNLTALVNGTTGAVLGOKAFDMYSGDVHMLVKFLKEIPGGALVLVASYDD 148
Db 61 PGTKNMDESKRLKFLSDLGSSYAKQLGFRDSWVFIGAKDLRGKSPFEQFLKNSPDNTNKYEGW 120
QY 149 PGTKNMDESKRLKFLSDLGSSYAKQLGFRDSWVFIGAKDLRGKSPFEQFLKNSPDNTNKYEGW 208
Db 121 PELLEMEGCMPPKPF 135
QY 209 PELLEMEGCMPPKPF 223

RESULT 3

ID W06547 standard; Protein; 135 AA.
AC W06547;
DT 13-MAR-1997 (first entry)
DE Human colon specific gene CSG4 polypeptide fragment.
KW Colon specific gene; CSG4; colon cancer; metastasis; diagnosis;
KW therapy; antibody; vaccine.
OS Homo sapiens.
PN W0639419-A1.
PD 12-DEC-1995.
PF 06-JUN-1995; U07289.
PR 06-JUN-1995; WO-007289.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen CA, Yu G;
DR WPI: 97-043054/04.
DR N-PSDB; T45883.
PT Human colon specific genes and their expression products - detection

PT of which, in non-colon tissue samples, can be used as indication of
PT colon cancer metastasis
PS Claim 8; Fig 4; 60pp; English.
CC Novel polypeptides (W06545-53) are encoded by cDNA clones (see also
CC T45880-92) corresponding to 13 human colon specific genes,
CC designated CSG1, CSG2, etc., that are primarily expressed in
CC tissues derived from the colon. Recombinant CSG polypeptides can
CC be produced in transformed host cells. They are useful diagnostic
CC markers for colon cancer and for colon cancer metastasis and can
CC also be used to screen for (ant)agonist cpds. of therapeutic or
CC diagnostic value. Antibodies raised against the colon-specific
CC polypeptides may be used to target colon cancer cells or as part
CC of a colon cancer vaccine.
SQ Sequence 135 AA;

Query Match 58.9%; Score 993; DB 1; Length 135;
Best Local Similarity 100.0%; Pred. No. 2.59e-84;
Matches 135; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 MSPVKNVGRGLNLTALVNGTTGAVLGOKAFDMYSGDVHMLVKFLKEIPGGALVLVASYDD 60
QY 89 MSPVKNVGRGLNLTALVNGTTGAVLGOKAFDMYSGDVHMLVKFLKEIPGGALVLVASYDD 148
Db 61 PGTKNMDESKRLKFLSDLGSSYAKQLGFRDSWVFIGAKDLRGKSPFEQFLKNSPDNTNKYEGW 120
QY 149 PGTKNMDESKRLKFLSDLGSSYAKQLGFRDSWVFIGAKDLRGKSPFEQFLKNSPDNTNKYEGW 208
Db 121 PELLEMEGCMPPKPF 135
QY 209 PELLEMEGCMPPKPF 223

RESULT 4

ID R97984 standard; Protein; 618 AA.
AC R97984;
DT 16-OCT-1996 (first entry)
DE DmORF1 potassium channel protein.
KW transmembrane helix; N-glycosylation site; pore-forming domain;
KW potassium-antagonist; drug screening; insecticide; cardiac disorder.
OS Drosophila melanogaster.
FH Key Location/Qualifiers
FT domain 6..27 /note= "Transmembrane domain M1"
FT modified_site 58..60 /note= "N-glycosylation site"
FT domain 95..111 /note= "Pore-forming domain H5-1"
FT domain 120..140 /note= "Transmembrane domain M2"
FT domain 171..191 /note= "Transmembrane domain M3"
FT domain 208..224 /note= "Pore-forming domain H5-2"
FT domain 242..263 /note= "Transmembrane domain M4"
FT WO9613520-A1.
PN 09-MAY-1996.
PD 25-OCT-1995; U14364.
PF 31-OCT-1994; US-33212.
PA (AMCY) AMERICAN CYANAMID CO.
PI Pausch MH, Price LA;
DR WPI: 96-239450/24.
DR N-PSDB; T33207.
PT Potassium channel genes from Drosophila melanogaster and
PT Caenorhabditis elegans - useful in assaying substances to determine
PT effects on cell growth, and in inhibiting nematode and insect pests
PS Claim 26; Page 37-39; 79pp; English.
CC This potassium channel sequence is encoded by the DmORF1 gene from
CC Drosophila melanogaster, and has 2 pore-forming domains between 4
CC hydrophobic transmembrane helix domains. Each pore-forming H5
CC domain contains a Y/F-G dipeptide motif required for potassium
CC selectivity. The protein contains a single N-terminal

CC asparagine-linked glycosylation site. The protein is predicted to
 CC span the membrane 4 times, with N- and C-termini within the cell,
 CC and the N-glycosylation site and H2 domains on the cell exterior,
 CC allowing permeation of the membrane by the pore-forming domains
 CC from the outside, a requirement for functional potassium channel
 CC formation. The protein may be expressed in a heterologous host
 CC cell to assay substances to determine effects on cell growth.
 CC Potassium-agonists or potassium-antagonists identified by this
 CC method may be used as insecticides or in therapy of cardiac
 CC disorders, etc.
 SQ Sequence 618 AA;

Query Match 6.4%; Score 108; DB 1; Length 618;
 Best Local Similarity 33.3%; Pred. No. 2.45e+00;
 Matches 20; Conservative 14; Mismatches 20; Indels 6; Gaps 6;
 Db 97 FFAFTVCTGVYGNISPTT-FAGRMIMAY-SVIGIPVNGILFAGL-GEVFG-RTPEAYI 152
 QY 64 YFAFKICSGAA-NVVGPTMCFEDRMIMSPVKNVGRGLNIALVNGTTGAVLGOKAFD-MY 121

RESULT 5
 ID W48720 standard; Protein; 993 AA.
 AC W48720;
 DT 19-AUG-1998 (first entry)
 DE Human mitochondrial isoleucyl-tRNA synthetase.
 KW Mitochondrial isoleucyl-tRNA synthetase; T-cell; pathogen.
 OS Homo sapiens.
 PN US5759833-A.
 PD 02-JUN-1998.
 PF 27-MAY-1994; 250852.
 PR 06-JUN-1995; US-468557.
 PR 27-MAY-1994; US-250852.
 PA (CANC-) CANCER INST JAPANESE FOUND CANCER.
 PA (CUBI-) CUBIST PHARM INC.
 PI Kranz JE, Schimmel PR, Shiba K;
 DR WPI; 98-3321142/29.
 DT N-PSDB; V18326.
 PR Human isoleucyl-tRNA synthetase DNA - useful for producing
 PT recombinant proteins, tester strains, etc.
 PS Claim 3; Columns 47-52; 46pp; English.
 CC The present sequence is that of a human mitochondrial isoleucyl-tRNA
 CC synthetase. The cDNA encoding for the mitochondrial isoleucyl-tRNA
 CC synthetase was isolated from a human T-cell cDNA library. The cDNA
 CC can be expressed using expression vectors to produce the corresponding
 CC recombinant protein. The protein can be used in assays to test
 CC substances known to inhibit the isoleucyl-tRNA synthetase or other tRNA
 CC synthetases of pathogenic organisms. Also, expression of the
 CC isoleucyl-tRNA synthetase in a tester strain can be useful for testing
 CC substances which are capable of inhibiting its activity.
 SQ Sequence 993 AA;

Query Match 6.0%; Score 102; DB 1; Length 993;
 Best Local Similarity 30.1%; Pred. No. 6.95e+00;
 Matches 22; Conservative 23; Mismatches 23; Indels 5; Gaps 5;
 Db 887 VIBPGLLFEIEMQSETSTSQ-LN-ELMMASESTLLAQEPREM-TADVIELKGFLLI 943
 QY 76 VVGPTMCFEDR-MIMSPVKNVGRGLNIALVNGTTGAVLGOKAFD-MYSGDMVHL-VKFLK 133
 Db 944 NLEGGDIRESSY 956
 QY 134 EIPGGLVLVASY 146

RESULT 6
 ID P81244 standard; protein; 187 AA.
 AC P81244;
 DT 16-NOV-1990 (first entry)
 DE Sequence of yellow tail fish pre-growth hormone
 KW pYGH1; somatotropin.
 OS Yellow tail fish.
 FH Key Location/Qualifiers

FT protein 3. 187
 FT /note="185 AA SQ, see CC"
 PN J63152985-A.
 PD 25-JUN-1988.
 PF 23-JUL-1987; 184083.
 PR 24-JUL-1986; JP-174385.
 PR 3-JUL-1987; JP-184083.
 PA (NAKA) Nakajima K.
 DR WPI; 88-216878/31.
 PT Escherichia coli plasmid used for growth of yellow tail fish -
 PT comprises plasmid pS14001 derived from ampicillin resistant gene
 PT obt'd. from pituitary tissue of yellow tail fish
 PS ; Fig 4; 7pp; Japanese.
 CC The E.coli plasmid comprises plasmid pS14001 derived from ampicillin
 CC resistant gene contg. E.coli plasmid pBR322 and yellow tail pre-growth
 CC hormone having cyclic double chain plasmid pYGH1 of 3851 BP, and yellow
 CC tail growth hormone of 185 and 187 (p81244) AA sequences contd. in pYGH1.
 CC The plasmids were introduced into E.coli host cell DH1. The plasmids
 CC are used for cultivation of young yellow tail, by E.coli and Bacillus
 CC subtilis having recombinant plasmid for yellow tail pre-growth hormone
 CC structural gene amplification.
 SQ Sequence 187 AA;

Query Match 5.6%; Score 95; DB 1; Length 187;
 Best Local Similarity 29.8%; Pred. No. 2.28e+01;
 Matches 14; Conservative 15; Mismatches 17; Indels 1; Gaps 1;
 Db 69 SSVLKLISYRLVSEWFSFSSFLSGSALRNQISPRSELKTEIQL 115
 QY 4 SGVLRLLALFAIVTWTFIRSYMSF-SMKTIRLPRLWLASPTKEIQV 49

RESULT 7
 ID W35714 standard; Protein; 476 AA.
 AC W35714;
 DT 27-MAR-1998 (first entry)
 DE Gentian flavonoid 3'-hydroxylase.
 KW Flavonoid 3'-hydroxylase; pigmentation; flower colour;
 KW transgenic plant; gentian.
 OS Gentiana triflora Pall. var japonica Hara.
 PN WO9732023-A1.
 PD 04-SEP-1997.
 PF 28-FEB-1997; AU0124.
 PR 01-MAR-1996; AU-008386.
 PA (FLOR-) FLORIGENE LTD.
 PI Bruggliera F, Holton TA, Michael MZ;
 DR N-PSDB; T94664.
 PT Novel flavonoid 3'-hydroxylase(s) from flowering plants - and
 PT corresponding DNA, used in the manipulation of pigmentation in
 PT plants
 PS Claim 23; Page 181-184; 234pp; English.
 CC This polypeptide comprises gentian flavonoid 3' hydroxylase
 CC (F3'H). Its amino acid sequence was deduced from a cDNA clone
 CC (see T94664) isolated from a petal cDNA library. F3'H acts on
 CC dihydrokaempferol to produce dihydroquercetin and on naringenin
 CC to produce eriodictyol. Isolated F3'H nucleic acids (see T94655-65)
 CC can be used to provide genetic constructs capable of modulating the
 CC hydroxylation of flavonoid compounds in a plant or in the cells of
 CC a plant, e.g. by de novo expression, over-expression, suppression,
 CC antisense inhibition and ribozyme activity. Modulation of F3'H is
 CC useful for the manipulation of pigmentation (flavonoids contribute
 CC to a range of colours from yellow to red to blue) in plants,
 CC especially flowering plants such as petunia, carnation,
 CC chrysanthemum, rose, snapdragon, tobacco, cornflower, pelargonium,
 CC lisianthus, gerbera, apple, iris, lily, african violet and morning
 CC glory.
 SQ Sequence 476 AA;

Query Match 5.6%; Score 95; DB 1; Length 476;
 Best Local Similarity 24.5%; Pred. No. 2.28e+01;
 Matches 23; Conservative 23; Mismatches 45; Indels 3; Gaps 2;

Db 88 LRKICSVHLEFSSKALDDFOVHRHEETICILIRAIASGGHAPVNLGKLGVCVTNALARVML 147
 QY 56 LKPCPANVFAFKICSGAANVVGPTMCFEDRMIPVKN--NVGRGLNALVNGTTGAVL 113
 Db 148 GRVPEGDGGEHNADEF-KSMVVEIWLAFN 180
 QY 114 GOKAFDMYSGVNDVHLVKFLKEIPGGALVLVASYD 147

RESULT 8

ID W62596 standard; Protein; 642 AA.
 AC W62596;
 DT 26-OCT-1998 (first entry)
 DE Human nuclear protein HEC.
 KW HEC gene; highly expressed in cancer; human; nuclear protein;
 KW mitosis; cell cycle; cell proliferation; malignancy; vaccine;
 KW antibody.
 OS Homo sapiens.
 FH Key Location/Qualifiers
 FT Modified_site 165
 FT 254..621
 FT /label= O-phosphorylated
 FT /note= "leucine heptad repeat region"
 PN W09827994-A1.
 PD 02-JUL-1998.
 PF 18-DEC-1997; U23385.
 PR 20-DEC-1996; US-033600.
 PA (TEXA) UNIV TEXAS SYSTEM.
 PI Chen P, Chen Y, Lee W, Riley DJ;
 DR WPI: 98-377401/32.
 DR N-PSDB; V38564.
 PT Nucleic acid encoding human nuclear protein HEC modulating mitosis -
 PT useful to, e.g control cell malignancies and other cell growth
 PT abnormalities at mitosis stage and to produce HEC protein and
 PT peptide(s)
 PS Claim 1; Page 63-65; 93pp; English.
 CC This protein comprises a novel human nuclear protein, designated
 CC HEC (highly expressed in cancer), that appears to be crucial for
 CC normal mitosis. Its amino acid sequence was deduced from an
 CC isolated B cell cDNA clone (see V38564). HEC localises to the
 CC nuclei of interphase cells and redistributes to centromeres during
 CC M phase. Ectopic expression of a mutant HEC containing only the
 CC leucine heptad repeats results in cells being unable to divide more
 CC than once. Inactivation of HEC results in disordered sister
 CC chromatid alignment and separation, and formation of non-viable
 CC cells with multiple, fragment micronuclei. HEC interacts through
 CC its leucine heptad regions with several proteins involved in
 CC mitosis, including nek2, sbl.8 and 2 different regulatory subunits
 CC of the 26S proteasome, Mss1 and p45. These properties suggest
 CC potential roles of HEC in modulating proteins important for spindle
 CC attachment to kinetochores, sister chromatic movement, and M phase
 CC progression. HEC nucleic acids and polypeptides are useful in
 CC compositions to control cell malignancies and other cell growth
 CC abnormalities at the cell mitosis stage. HEC peptides are useful
 CC in vaccines and may be administered to cells to disrupt chromatid
 CC separation and so modulate cell cycle progression (claimed). HEC
 CC specific antibodies are useful in purifying native or recombinant
 CC HEC and to detect HEC protein/peptides in samples (claimed)
 CC immunodetection kits are provided). They can also be administered
 CC to disorder sister chromatid alignment and separation in an
 CC interphase cell, so disrupting mitosis (claimed).
 SQ Sequence 642 AA;

Query Match 5.6%; Score 94; DB 1; Length 642;
 Best Local Similarity 40.5%; Pred. No. 2.70e+01;
 Matches 17; Conservative 9; Mismatches 15; Indels 1; Gaps 1;
 Db 122 VKDFLKIFTLGFLCPSYELDPDKFEEVPRFKDLGYPPA 163
 QY 129 VK-FLKEIPGALVLVASYDDPQTKMNDSRKLFSLDGSSYA 169
 RESULT 9

ID W80950 standard; Protein; 642 AA.
 AC W80950;
 DT 16-FEB-1999 (first entry)
 DE Amino acid sequence of the human nuclear protein.
 KW Human nuclear protein; HEC; cell growth abnormality; hybridisation;
 KW amplification; antibody; immunoassay; malignancy; mitosis; antigen;
 KW immunoprecipitation; immunisation; vaccine.
 OS Homo sapiens.
 PN W09845433-A1.
 PD 15-OCT-1998.
 PF 03-APR-1998; U06727.
 PR 04-APR-1997; US-042609.
 PA (TEXA) UNIV TEXAS SYSTEM.
 PI Allred DC, Channess GC, Clark GM, Hilsenbeck SG,
 PI Osborne CK;
 DR WPI: 98-594481/50.
 DR N-PSDB; V68588.
 PT New nucleic acid encoding human highly expressed in cancer nuclear
 PT protein - used for diagnosis and for modulation of the cell cycle to
 PT control malignant and other cell growth abnormalities
 PS Claim 2; Pages 58-61; 75pp; English.
 CC This is the amino acid sequence of a human nuclear protein (HEC) used
 CC in the method of the invention to diagnose cell growth abnormalities.
 CC fragments of HEC are used to detect HEC-encoding nucleic acid, in
 CC usual hybridisation or amplification assays, while antibodies are used
 CC in conventional immunoassays to detect HEC or peptides. HEC is used to
 CC modulate cell cycle progression (by disrupting chromatid separation)
 CC and antibodies are used to disorder sister chromatid alignment and
 CC separation in interphase cells, disrupting mitosis, particularly for
 CC control of malignancy and other cell growth abnormalities at the
 CC mitosis stage. Antibodies can also be used for protein purification, to
 CC isolate sequences encoding HEC or related proteins; to study HEC
 CC distribution in cells and to isolate antigens by immunoprecipitation.
 CC HEC may also be used in vaccines and antibodies for passive
 CC immunisation.
 SQ Sequence 642 AA;
 Query Match 5.6%; Score 94; DB 1; Length 642;
 Best Local Similarity 40.5%; Pred. No. 2.70e+01;
 Matches 17; Conservative 9; Mismatches 15; Indels 1; Gaps 1;
 Db 122 VKDFLKIFTLGFLCPSYELDPDKFEEVPRFKDLGYPPA 163
 QY 129 VK-FLKEIPGALVLVASYDDPQTKMNDSRKLFSLDGSSYA 169
 RESULT 10
 ID R42381 standard; Protein; 146 AA.
 AC R42381;
 DT 19-APR-1994 (first entry)
 DE Haemin-binding protein produced as fusion protein in pGCH5.
 KW Haemophilus somnus; immunogenic; haemolysin; LppB; LppC;
 KW thromboembolic meningencephalitis; septicaemia; arthritis;
 KW pneumonia; lktA gene; Pasteurella haemolytica.
 OS Haemophilus somnus.
 PN W09321323-A.
 PD 28-OCT-1993.
 PF 05-APR-1993; CA0135.
 PR 09-APR-1992; US-865050.
 PR 04-JUN-1992; US-893424.
 PR 04-JUN-1992; US-893426.
 PR 29-MAR-1993; US-038287.
 PR 29-MAR-1993; US-038288.
 PR 29-MAR-1993; US-038719.
 PR (UYSA-) UNIV SASKATCHEWAN.
 PA Harland RJ, Pfeiffer CG, Pontarollo RA, Potter AA;
 PI Rioux C, Theisen M;
 DR WPI: 93-351733/44.
 DR N-PSDB; Q51082.
 PT Haemophilus somnus immunogenic proteins used in vaccines -
 PT selected from haemin-binding protein, haemolysin, LppB and LppC,
 PT and corresp. DNA
 PS Claim 1; Fig 6; 119pp; English.

KW	thromboembolic meningoenophthalitis; septicaemia; arthritis;
KW	pneumonia; haemin-binding protein.
OS	Haemophilus somnus.
PN	WQ9321323-A.
PD	28-OCT-1993.
PF	05-APR-1993; CA0135.
PR	03-APR-1992; US-865050.
PR	04-JUN-1992; US-893424.
PR	04-JUN-1992; US-893426.
PR	29-MAR-1993; US-038287.

PR 29-MAR-1993; US-038719.
PA (UYSA-) UNIV SASKATCHEWAN.
PI Harland RJ, Pfeiffer CG, Pontarollo RA, Potter AA;

DR WPI; 93-351733/44.
DR N-PSDB; Q51080.
PT Haemophilus somnus immunogenic proteins used in vaccines -

CC A genomic cosmid library of *Haemophilus somnus* HS25 DNA was screened

activity. Positive clones were subcloned various times, resulting in plasmid pRAP50L, which binds haemin but is not haemolytic. The clone was sequenced and was found to contain several open reading frames, potentially encoding 8 proteins. The haemin binding protein (encoded by the hmb gene) was encoded by ORF1. The protein can be used in vaccines for preventing or treating *H. somnus* infections, which cause thromboembolic meningo-encephalitis, septicæmia,

[illegible]

RESULT	13	
ID	R10967	standard; Protein; 191 AA.
AC	R10967;	
DT	16-APR-1991	(first entry)
DE	Bovine Somatotropin analogue #20.	
DE	bovine somatotropin; bST; dairy cow; helical stability.	
KW	Bos taurus.	
OS	Key	Location/Qualifiers
FH	misc_difference	126
FT	/label= Ala, Cys, Leu, Gln, His, Lys, Val, Ile,	
FT	Phe, Tyr, Trp, Thr, Gly, Ser, ASP, Asn,	
FT	Pro, Arg	
FT	WO9100870-A.	
PN	24-JAN-1991.	
PD		

PR 10-JUL-1989; US-377926.
PA (UPJO) UPJOHN CO.
PI Lehman SR, Havel HA, Plaisted SM, Brems DN;

WFL; 91-0313/07.
 DR Animal somatotropin analogues - have aminoacid changes at
 PT residues 96 to 133 to reduce hydrophobicity or helical stability
 PT of somatotropin
 PT

The analogue-BST has a lower propensity of the partially denatured hydrophobicity and helical stability in the region of the mutation.

protein for aggregation and precipitation during processing. It also has enhanced bioactivity and can be used for enhancing the growth of an animal, e.g. beef cattle or increasing milk production in dairy cows. The mutation was introduced by standard site-

Search completed: Wed May 31 11:03:33 2000
Job time : 12 secs.

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	1198	98.1	1299	1	V63516	Consensus sequence
2	941	77.1	955	1	V63515	Full length sequen
3	515.6	42.2	548	1	V16671	Polynucleotide seq
4	514	42.1	548	1	T45883	Human colon specif
c	286	23.4	301	1	V63513	Clone g2197992 EST
5	286	23.4	301	1	V63513	Clone g2197992 EST
6	269.4	22.1	287	1	V63508	Clone 3210629 EST
7	263	21.5	263	1	V63509	Clone 1281865 EST
8	251	20.6	251	1	V63511	Clone 1735382 EST
9	228	18.7	228	1	V63510	Clone 1807758 EST
10	209	17.1	233	1	V63507	Clone 775762 EST c
11	196.4	16.1	250	1	V63506	Clone 2645837 EST
12	179.2	14.7	187	1	V63512	Clone 2128334 EST
13	165.6	13.6	250	1	V63505	Clone 1338704 EST
14	163.2	13.4	273	1	V63504	Clone 775133 EST c
c	114	9.3	123	1	V63514	Clone g2409650 EST
15	114	9.3	123	1	V63514	Clone g2409650 EST
16	39.2	3.2	2358	1	Q07000	Coding part of vec
17	39.2	3.2	2440	1	Q06469	Sequence encoding
18	39.2	3.2	3086	1	O06470	Sequence encoding
19	38	3.1	6803	1	T85473	Genomic hNET seque
20	38	3.1	6803	1	V16305	Genomic DNA encodi
21	35.8	2.9	3475	1	Q21453	Sequence encoding
22	34.6	2.8	1561	1	Q25420	Encodes human live
23	34.4	2.8	2363	1	O58126	Phytase gene. Comp
24	34.4	2.8	2379	1	O56944	A. niger phytase g
c	34	2.8	1521	1	V01503	Human p53 tumour s
25	34	2.8	1521	1	V01503	Human secreted pro
26	33.8	2.8	1486	1	V59523	Serum opacity fact
c	33.8	2.8	3240	1	Q83181	OPDE 45 kDa subuni
c	33.6	2.8	2085	1	K81504	HSV-2 strain SB5 C
c	33.4	2.7	2341	1	V62155	HSV-2 strain SB5 C
30	33.4	2.7	2694	1	V62145	HSV-2 strain SB5 C
c	33.4	2.7	11213	1	V62176	Chromosomal phytas
31	33.4	2.7	1344	1	Q11175	Phytase gene. Phyt
32	33	2.7	1404	1	Q13878	Aspergillus ficuum
33	33	2.7	1404	1	T65136	Aspergillus ficuum
34	33	2.7	1404	1	T65136	Aspergillus ficuum

Db	1	ATGAGTCCCTGTGAAAACAAATGGGCAGAGCCATAACATGCCCTGGTGAATGAAC	60
Qy	546	acgggagctgtcgtggacagaggcatttgacatgtactctggagatgttatgcaccta	605
Db	61	ACGGGAGCTGTCTGGACAGAGGCAATTTGACATGTACTCTGGAGATCTTATGCACCTA	120
Qy	606	gtgaatatcctttaaagaataatccggggggtgcactgtgtgtgtgtgtgtcctctcaacacat	665
Db	121	GTGAAATTCCTTTAAGAAATTCGGGGGGTGCACTGGTCTGGTGGCCCTCTACGACGAT	180
Qy	666	ccagggaacaaatgaacgatgaaacaggaacactctctgacttgggaagttccctac	725
Db	181	CCAGGGACCAAAATGAAACGATGAAGACGAAACTCTCTCTGACTTGGGAGTTCCTTAC	240
Qy	726	qcaaaacaacttggggtccgggacagctgggtcttcataggagccaaagacctcaggggt	785
Db	241	GCAAAACAACCTGGCTTCCGGGACAGCTGGGTCTTCATAGGACCAAGACCTCAGGGGT	300
Qy	786	aaagccctttgagcagttctttaaagaacagccacagacacaaatacagagggatg	845
Db	301	AAAAGCCCTTTGAGCAGTCTTTAAAGAACAGCCACAGACACAAACAAATACGAGGGATGG	360
Qy	846	ccgagactgtgagatgaaggctcatgcccccgaagccattttaggtggctgtggc	905
Db	361	CCAGAGCTGCTGGAGATGGAGGGCTGCATGCCCCCGAAGCCATTTTAGGGTGGCTGTGGC	420
Qy	906	tcttctccagccaggggccctgaagaagctcctctgctgacttaggagtcagagcccgagg	965
Db	421	TCTTCTCACCAGGGGCTTGAAGAGYTCCCTGCCCTGCATTAGGAGTCANAGCCCGGCA	480
Qy	966	gggctgaggaggagcaggggggtgctgcgtggaaggtgctgcagaggtccttgcacgtg	1025
Db	480	GGCTGNAGGAGGAGACAGCGGGTGTGCTGTGAAGGTGCTGCAGG - CTTTGCACGTGTG	538

Qy	1026	tgctgcgcct	1035
Db	539	TGTCGCCT	548
RESULT	4		
ID	T45883	standard; cDNA; 548 BP.	
AC	T45883;		
DT	13-MAR-1997	(first entry)	
DE	Human colon specific gene CSG4 cDNA partial clone.		
KW	Colon specific gene; CSG4; colon cancer; metastasis; diagnosis;		
KW	gene therapy; ss.		
OS	Homo sapiens.		
Key	Location/Qualifiers		
FT	cds	1..408	
FT		/*tag= a	
PN	W09639419-A1.		
PD	12-DEC-1996.		
PF	06-JUN-1995; U07289.		
PR	06-JUN-1995; WO-U07289.		
PA	(HUMA-) HUMAN GENOME SCI INC.		
PI	Rosen CA, Yu G;		
DR	WPI: 97-043054/04.		
DR	P-PSDB: W06547.		
PT	Human colon specific genes and their expression products - detection		
PT	of which, in non-colon tissue samples, can be used as indication of		
PT	colon cancer metastasis		
PS	Claim 1: Fig 4: 60pp: English.		
CC	13 cDNA clones (T45880-92), most of them partial clones, correspond		
CC	to human colon specific genes, designated CSG1, CSG2, etc., that		
CC	are primarily expressed in tissues derived from the colon. CSG7		
CC	and CSG10 show reduced expression in colon cancer cells as compared		
CC	to that in normal cells; the remaining genes are overexpressed in		
CC	colon cancer. The partial cDNA sequences can be used to isolate		
CC	full-length clones and genomic clones including the complete gene.		
CC	CSG nucleic acids can be used to produce CSG polypeptides (see also		
CC	W06545-53) in transformed host cells, as probes to detect disorders		

Query Match 42.2%; Score 515.6; DB 1; Length 548;
Best Local Similarity 98.2%; Pred. NO. 1e-141;
Matches 540; Conservative 1; Mismatches 7; Indels 2; Gaps
Oy 486 atdaggctctgtgaaacaaatgtggcgacagagccttaaacatcgcctcggtgaatggaacc 545

CC of the colon, partic. colon cancer and colon cancer metastasis, and
 CC in gene therapy.
 SQ Sequence 548 BP; 137 A; 129 C; 172 G; 107 T;

Query Match	42.1%	Score 514;	DB 1;	Length 548;
Best Local Similarity	98.0%;	Pred. No. 3e-141;		
Matches 539;	Conservative 1;	Mismatches 8;	Indels 2;	Gaps
QY	486	atagtcctgtgaaaaacaatgtggcgagaggcctaaacatgcctcgtgtgaatggaacc	545	
DB				
	1	ATGAGTCCTGTGAAAAACAATGTGGCAGAGGCTTAACATCCCTGTGTGAATGGAACC	60	
QY	546	acggggagctgtctgggcacagaaggcatttgacatgtactctctggagatgttatgcacct	605	
DB				
	61	ACGGGAGCTGTCTGGGACAGAGGCATTTGACATGTACTCTGGAGATGTTATGCACCTA	120	
QY	606	gtgaatactcttaagaaatcccggggggtgcactggtgctggtgctcctctacgacgat	665	
DB				
	121	GTGAATAATTCCTTAAGAAATTCGGGGGTTGCATGCTGTGTGGCTCTCTACGACGAT	180	
QY	666	ccaggaccacaaatgaacgatgaacgagaaaactctctgaactggggagttcctac	725	
DB				
	181	CCAGGACCAAAATGAACGATGAAGCAGGAAACTCTTCTGACTTGGGAGTTCTCTAC	240	
QY	726	gcaaaacaactgggctccggggacagctgggtcttcattaggagccaaaacacctcagggt	785	
DB				
	241	GCAAAACAATGGGCTTCCGGGACAGCTGGGTCTTCATAGGAGCCAAAGACCTCAGGGT	300	
QY	786	aaagaccctttgagcagttcttaagaacagcccagacacaaaaatacagaggatgg	845	
DB				
	301	AAAGCCCCCTTGAGCAGTTCTTTAAAGACAGCCCCAGACAAACAAAATCAGAGGATGG	360	
QY	846	ccagagctgtcgagatggaggctgcattgccccgaagccattttagggtggctgtgac	905	
DB				
	361	CCAGAGCTGTGGAGATGGAGGCTGCATGCCCCCAGGCCATTTTAGGGTGGCTGTGGC	420	
QY	906	tcttctcagccaggggcttgaaagaagctctctgcctgactttaggagtcagagcccgag	965	
DB				
	421	TCNTTCCTCAAGCCAGGGGCTTGAAGAAGYTCCTGCCGTGCATTAGGAGTCANAGCCGCCA	480	
QY	966	gggctgagagagagagcagggggtactcgtggaaggtgctcaggtccttgcacgtg	1021	
DB				
	480	GGCTGNAGGAGGAGACAGGGGGTCTCGGTGGAGAGTGTCTGCAGG-CCTTGCACGCTG	538	
QY	1026	tgtcgcgctt	1035	
DB				
	539	TGTCGGGCT	548	

RESULT	5	
V63513/C		
V63513	standard; cDNA; 301 BP.	
AC	V63513;	
DT	28-JAN-1999 (first entry)	
DE	Clone g2197992 EST corresponding to CS141 sequence.	
DE	EST sequence; CS141; gastrointestinal tract; cancer; Barrett's esophagus;	
KW	gastric ulcer; gastritis; leiomyoma; polyps; Crohn's disease;	
KW	ulcerative colitis; pancreatitis; ss.	
OS	Synthetic.	
OS	Homo sapiens.	
PN	W09844133-Al.	
PD	08-OCT-1998.	
PF	31-MAR-1998: U06337.	
PR	31-MAR-1997: US-828489.	
PA	(ABRO) ABBOTT LAB.	
PI	Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,	
PI	Gordon J, Granados EN, Hayden M, Hodges SC, Klass MR,	
PI	Kratovich JD, Roberts-Rapp L, Russell JC, Stroupe SD;	
PI	WPI: 98-568280/48.	
DR	New gastrointestinal tract specific polynucleotides, CS141 - used to	
PPT	develop products for the diagnosis and treatment of e.g. cancers,	
PT	gastric ulcer, gastritis, Crohn's disease, ulcerative colitis or	
PT		

pancreatitis
Claim 1; Page 92; 116pp; English.
CC V63504-15 represent a set of contiguous and partially overlapping EST
CC sequences designated CS141. The sequences are isolated from a cDNA
CC library made from gastrointestinal tract tumour and normal tissues.
CC The CS141 gene is useful as a marker for gastrointestinal tract
CC disorders. The methods and products can be used in detecting, or
CC diagnosing, staging, monitoring, prognosticating, preventing, or
CC determining the predisposition to diseases and conditions of the
CC gastrointestinal tract, such as gastrointestinal tract cancers,
CC Barrett's esophagus, gastric ulcer, gastritis, leiomyoma, polyps,
CC Crohn's disease, ulcerative colitis, pancreatitis.
SQ Sequence 301 BP; 63 A; 92 C; 77 G; 69 T;

Query Match	23.4%	Score 286;	DB 1;	Length 301;
Best Local Similarity	100.0%;	Pred. No. 1.6e-74;		
Matches 286;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	935	cctgcctgaacttaggagtcagagcccgccaggggctgaggaggagcaggggggtgctg	994
Db	301	CCTGCTCTACATTAGGAGTTCAGAGCCGCGCAGGGGCTGAGGAGGAGGAGCGGGGTGCTG	242
Qy	995	cgtaggaaggctgcacaggtcccttcacagcgtgtgtgcgcctctcctcctcggaacagaa	1054
Db	241	CGTGGAAAGGTGCTGCAGGTCCTTCACGCTGTGTGCGGCTCTCTCTCGGAACAGAA	182
Qy	1055	ccctccacacagacatcctaccggaagacagcctcagagggtccctctgtgaaccagct	1114
Db	181	CCCTTCCACAGCACATCCTATCCCGGAAGACAGGCTCAGAGGGTCTTCTTGGAAACACAGCT	122
Qy	1115	gtctgtgtgagagaatgggtgtcttcgtccaggagactgctgacggctggctcctgaggaaag	1174
Db	121	GTCGTGTGGAGAGATGGGTGCTTTTGTCTCAGGGACTGCTGACGGCTGGTCTCTGAGGAGG	62
Qy	1175	acaaactgccagacttgagcccaattaaatttttttgcgtggt	1220
Db	61	ACAAACTGCCCAGACTTGAGCCCAATTAATTTTATTTTTCCTGCTGCT	16

RESULT	6
V63508	V63508 standard; cDNA; 287 BP.
AC	V63508;
AT	28-JAN-1999 (first entry)
DE	Clone 3210629 EST corresponding to CS141 sequence.
DE	EST sequence; CS141; gastrointestinal tract; cancer; Barrett's esophagus;
KW	gastric ulcer; gastritis; leiomyoma; polyps; Crohn's disease;
KW	ulcerative colitis; pancreatitis; ss.
OS	Synthetic.
OS	Homo sapiens.
PN	WO98441133-A1.
PD	08-OCT-1998.
PF	31-MAR-1998; U06337.
PR	31-MAR-1997; US-828489.
PR	(ABBO) ABBOTT LAB.
PI	Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
PI	Gordon J, Granados EN, Hayden M, Hodges SC, Klass MR,
PI	Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;
DR	WPI: 98-568280/48.
PT	New gastrointestinal tract specific polynucleotides, CS141 - used to
PT	develop products for the diagnosis and treatment of e.g. cancers,
PT	gastric ulcer, gastritis, Crohn's disease, ulcerative colitis or
PT	pancreatitis
PT	Claim 1; Page 90; 116pp; English.
PS	V63504-15 represent a set of contiguous and partially overlapping EST
CC	sequences designated CS141. The sequences are isolated from a cDNA
CC	library made from gastrointestinal tract tumour and normal tissues.
CC	The CS141 gene is useful as a marker for gastrointestinal tract
CC	disorders. The methods and products can be used in detecting,
CC	diagnosing, staging, monitoring, prognosticating, preventing, or
CC	determining the predisposition to diseases and conditions of the
CC	gastrointestinal tract, such as gastrointestinal tract cancers.

Db	1	GATGTTATTGGAAGCTACATGAGCTTCAGCATGAAACCATCGTCTGCCAGTGGCT	60	
Qy	341	ggctcgccaccacaaggagatccaggttaaaagtacaagtggcctcatcaagccctg	400	
Db	61	GGCTCGCCACCACAGAGATCCAGGTTAAAAGTACAAGTGTGGCTTCATCAAGCCCTG	120	
Qy	401	cccagccaaactacttgcgtttataaatcgcagtggggcgcacaacgctcgtggccctac	460	
Db	121	CCAGGCCAACTACTTTTCGCTTTAAAAATCTGCAGTGGGGCGGCAACGTCGTGGCCCTAC	180	
Qy	461	tatgtctttgaagaccgcatatcatgactcctgtgaaaaaactgtgggcagaggcct	520	
Db	181	TATGTGCTTTGAAGACCGCATGATCATGAGTCTCTGTGAATAACAATGTGGCAGAGCCT	240	
Qy	521	aaacatcgccctggtgaatgaa 543		
Db	241	AAACATCGCCCTGGTGAATGGAA 263		
RESULT	8			
V63511	ID	V63511 standard; cDNA; 251 bp.		
AC	V63511;			
DT	28-JAN-1999	(first entry)		
DE	Clone 1735382 EST corresponding to CS141 sequence.			
KW	EST sequences; CS141; gastrointestinal tract; cancer; Barret's esophagus;			
KW	gastric ulcer; gastritis; leiomyoma; polyps; Crohn's disease;			
KW	ulcerative colitis; pancreatitis; ss.			
OS	Synthetic.			
OS	Homo sapiens.			
PN	WO9844133-A1.			
PD	08-OCT-1998.			
PF	31-MAR-1998;	U06337.		
PR	31-MAR-1997;	US-828489.		
PA	(ABBO) ABBOTT LAB.			
PI	Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,			
PI	Gordon J, Granados EN, Hayden M, Hughes SC, Klass MR,			
PI	Kratovich JD, Roberts-Rapp L, Russell JC, Stroupe SD;			
PR	WPI: 98-568280/48.			
PT	New gastrointestinal tract specific polynucleotides, CS141 - used to			
PT	develop products for the diagnosis and treatment of e.g. cancers,			
PT	gastric ulcer, gastritis, Crohn's disease, ulcerative colitis or			
PT	pancreatitis			
PS	Claim 1; Page 91; 116pp; English.			
CC	V63504-15 represent a set of contiguous and partially overlapping EST			
CC	sequences designated CS141. The sequences are isolated from a cDNA			
CC	library made from gastrointestinal tract tumour and normal tissues.			
CC	The CS141 gene is useful as a marker for gastrointestinal tract			
CC	disorders. The methods and products can be used in detecting,			
CC	diagnosing, staging, monitoring, prognosticating, preventing, or			
CC	determining the predisposition to diseases and conditions of the			
CC	gastrointestinal tract, such as gastrointestinal tract cancers,			
CC	Barret's esophagus, gastric ulcer, gastritis, leiomyoma, polyps,			
CC	Crohn's disease, ulcerative colitis, pancreatitis.			
SQ	Sequence	251 bp; 69 A; 62 C; 76 G; 44 T;		
Query Match	20.6%;	Score 251; DB 1: Length 251;		
Best Local Similarity	100.0%;	Pred. No. 2.7e-64;		
Matches	251; Conservative	0; Mismatches	0; Indels	0; Gaps
Qy	629	gggggggtgactggtgctggcgctctacgacgattccaggaccacaaatgaacgatga	688	
Db	1	GGGGGGTGCATGTGTGTGGCTCTACGAGATCCAGGACCAAAATGAACGATGA	60	
Qy	689	aagcagagaaactcttctgactggggagttcctacgcaaaacactgggcttcggga	748	
Db	61	AAGCAGGAAACTCTTCTGACTTTGGGAGTTCTACGCCAAACAACTGGGCTCCCGGA	120	
Qy	749	cagctgggtcttcataaggagccaaagacctcagggttaaaagccctttgagcagttctt	808	
Db	121	CAGCTGGGTCTTCATAGSAGCAAGACCTCAGGGGTAAAGGCCCTTTGAGGTCTTT	180	

CC	Gastrointestinal tract, such as gastrointestinal tract cancers,
CC	Barret's esophagus, gastric ulcer, gastritis, leiomyoma, polyps,
CC	Crohn's disease ulcerative colitis, pancreatitis.
SQ	Sequence 187 BP; 35 A; 42 C; 74 G; 33 T;
Query Match	14.7%; Score 179.2; DB 1; Length 187;
Best Local Similarity	96.8%; Prod. No. 2.6e+43;
Matches 181; Conservative 0;	Mismatches 0; Gaps 0;
QY	839 gggatgccagagtctgtggagatggaaggcgtcatgcgcccgaaaccatttagtggtg 898
Db	1 GGATGCCAGAGTCTGTGAGATGGAGGCTGCATGCCCGAAGCCATTTTAGSGTG 60
QY	ctgtggcttcttcagccaggggcctgaagaacctcctgacttaggaatcaaac 958
Db	61 CTGTGGCTCTCCTCAGCCAGGGGCGCTGAAGAAGCTCCTGCCTGACTTAGGAGCTAGAC 120.

Db 121 CCGCAGGGGCTGAGGAGGAGGAGNGGGTGCTGCCGTGAAGGTGCTGCACAGTCCCTG 180

Qy 1019 cagcgtg 1025

Db	181	AAAGNNG	187
	RESULT	13	
	V63505		
	ID	V63505 standard; cDNA; 250 BP.	
	AC	V63505;	
	DT	28-JAN-1999 (first entry)	
	DE	Clone 1338704 EST corresponding to CS141 sequence.	
	KE	EST sequence; CS141; gastrointestinal tract; cancer; Barret's esophagus;	
	KW	gastric ulcer; gastritis; leiomyoma; polyps; Crohn's disease;	
	KW	ulcerative colitis; pancreatitis; ss.	
	OS	Synthetic.	
	OS	Homo sapiens.	
	PN	W09844133-Al.	

31-MAR-1997 : US-828489.
PR (ABBO) ABBOTT LAB.
PA Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
PI Gordon J, Granados EN, Hayden M, Hodges SC, Klass MR,
PI Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;
MPI: 98-568280/46.
DR New gastrointestinal tract specific polynucleotides, CSI41 - used to
PT develop products for the diagnosis and treatment of e.g. cancers,
PT gastric ulcer, gastritis, Crohn's disease, ulcerative colitis or
PT

PS Claim 1; Page 89; 116pp; English

CC V63504-15 represent a set of contiguous and partially overlapping EST

CC sequences designated CS141. The sequences are isolated from a cDNA

CC library made from gastrointestinal tract tumour and normal tissues.

CC The CS141 gene is useful as a marker for gastrointestinal tract

CC disorders. The methods and products can be used in detecting,

CC diagnosing, staging, monitoring, prognosticating, preventing, or

CC determining the predisposition to diseases and conditions of the

CC gastrointestinal tract, such as gastrointestinal tract cancers,

CC Barrett's esophagus, gastric ulcer, gastritis, leiomyoma, polyps,

CC Crohn's disease, ulcerative colitis, pancreatitis.

CC Sequence 250 BP; 44 A; 80 C; 84 G; 38 T;

SQ

Query Match 13.6%; Score 165.6; DB 1; Length 250;

[illegible]

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 31, 2000, 17:30:02 ; Search time 57.43 Seconds
(without alignments)
2763.569 Million cell updates/sec

Title: US-09-186-342-1
Perfect score: 1221
Sequence: 1 cgccggcagggtgcattc.....aaattttattttgctggta 1221

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	515.6	42.2	548	2	US-08-469-667-6
2	515.6	42.2	548	6	PCT-US95-07289-6
3	46.6	3.8	7218	1	US-08-232-463-14
4	38	3.1	6803	5	US-08-665-259-19
5	35.8	2.9	3475	2	US-07-960-389-1
6	34.4	2.8	2363	2	US-07-923-724-7
7	34.4	2.8	2363	3	US-08-609-426A-7
8	34.4	2.8	2379	3	US-08-374-652C-1
9	33.8	2.8	3240	3	US-08-294-189-2
10	33.6	2.8	2085	3	US-08-283-917-8
11	33.6	2.8	2085	3	US-08-961-716-8
12	33	2.7	1404	1	US-08-151-574-33
13	33	2.7	1404	1	US-08-146-424-19
14	33	2.7	1404	3	US-08-693-709-1
15	33	2.7	1404	3	US-08-419-448-33
16	33	2.7	6756	1	US-08-151-574-31
17	33	2.7	6756	3	US-08-419-448-31
18	33	2.7	31571	1	US-08-323-443B-1
19	32.6	2.7	579	2	US-08-190-560-3
20	32.6	2.7	579	2	US-08-469-277-3
21	32.6	2.7	579	3	US-08-468-946-3
22	32.6	2.7	579	4	US-08-468-942-3
23	32	2.6	2923	2	US-08-377-292-6
24	32	2.6	2923	3	US-07-989-847-7
25	32	2.6	2923	7	5187076-5
26	32	2.6	5057	3	US-08-365-486A-12
27	32	2.6	5108	1	US-07-642-002-1

28	31.8	2.6	340	5	US-08-888-077A-31	Sequence 31, Appl
29	31.6	2.6	1947	1	US-07-951-715A-3	Sequence 3, Appl
30	31.6	2.6	1947	3	US-08-459-448A-3	Sequence 3, Appl
31	31.6	2.6	1947	5	US-08-459-595A-3	Sequence 3, Appl
32	31.6	2.6	3468	1	US-07-951-715A-4	Sequence 4, Appl
33	31.6	2.6	3468	1	US-07-951-715A-8	Sequence 8, Appl
34	31.6	2.6	3468	3	US-08-459-448A-4	Sequence 4, Appl
35	31.6	2.6	3468	3	US-08-459-448A-8	Sequence 8, Appl
36	31.6	2.6	3468	5	US-08-459-595A-4	Sequence 4, Appl
37	31.6	2.6	3468	5	US-08-459-595A-8	Sequence 8, Appl
38	31.6	2.6	3546	1	US-07-951-715A-10	Sequence 10, Appl
39	31.6	2.6	3546	1	US-07-951-715A-12	Sequence 12, Appl
40	31.6	2.6	3546	1	US-07-951-715A-14	Sequence 14, Appl
41	31.6	2.6	3546	1	US-07-951-715A-27	Sequence 27, Appl
42	31.6	2.6	3546	3	US-08-459-448A-10	Sequence 10, Appl
43	31.6	2.6	3546	3	US-08-459-448A-12	Sequence 12, Appl
44	31.6	2.6	3546	3	US-08-459-448A-14	Sequence 14, Appl
45	31.6	2.6	3546	3	US-08-459-448A-27	Sequence 27, Appl

ALIGNMENTS

RESULT 1
US-08-469-667-6
; Sequence 6, Application US/08469667
; Patent No. 573748
; GENERAL INFORMATION:
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Rosen, Craig
; TITLE OF INVENTION: Colon Specific Genes and Proteins
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
; ADDRESSEE: Stewart & Olstein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: NJ
; COUNTRY: USA
; ZIP: 07068-1739
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA: US/08/469,667
; APPLICATION NUMBER: US/08/469,667
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferraro, Gregory D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-435
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 548 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..405
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1..405
US-08-469-667-6

Query Match 42.2%; Score 515.6; DB 2; Length 548;

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 548 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..405
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1..405
; PCT-US95-07289-6

Query Match 42.2%; Score 515.6; DB 6; Length 548;
Best Local Similarity 98.2%; Pred. No. 3e-142;
Matches 540; Conservative 1; Mismatches 7; Indels 2; Gaps 2;

QY 486 atgagtcctgtgaaacaaatgtggcagagcctaaacatcgccctgggtgaatggaacc 545
Db 1 ATGAGTCCTGTGAAACAAATGTGGCAGAGCCCTAAACATCGCCCTGGTGAATGGAACC 60
QY 546 acgggagctgtgctgggacagaagcatttgacatgtactctgagatgttatgcaccta 605
Db 61 ACGGAGCTGTGCTGGACAGAGGCAATTTGACATGTACTCTGAGATGTTATGCACCTA 120
QY 606 gtgaattccctaaagaattccggggggtgacctgtgtgtgctgacctctacacgat 665
Db 121 GTGAAATTCCTTAAAGAAATTCGGGGGGTGCACCTGGTGGCTCCTACGACGAT 180
QY 666 ccagggaacaaatgaacgatgaagcagaaactcttctgagctgagagcccgagctctac 725
Db 181 CCAGGGACCAAAATGAACGATGAAGAGGAGGAACTCTTCTGACTTGGGGAGTTCTCTAC 240
QY 726 gcaaaacaactgggctccgggacagctgggtctctcatagagcgaacacccctcagggt 785
Db 241 GCAAAACAACCTGGGCTTCGGGGACAGCTGGGTCTTCATAGGAGCCCAAGACCTCAGGGT 300
QY 786 aaaagcccttgagcagctcttaaaagacagccagacacacacacacacacacacacac 845
Db 301 AAAAGCCCTTTGAGCAGTTCTTAAAGAACAGCCACAGACACAAACAAATACGAGGATGG 360
QY 846 ccagagctgtgagatgagctgacatgacatgacatgacatgacatgacatgacatgacat 905
Db 361 CCAGAGCTGTGAGATGAGGAGGCTGCATGCCGCCAGGCAATTTAGGGTGGCTGTGGC 420
QY 906 tcttctcagccaggggctgaaagagctgctgctgctgctgctgctgctgctgctgctgctg 965
Db 421 TCTTCTCCTCAGCCAGGGGCTGAAAGAGYTCCTGCTGCTCATTAGGAGTCANAGCCCGCA - 480
QY 966 gggctgagagagagagcaggggtgctgctgagagagagagagagagagagagagagagag 1025
Db 480 GGCTGNAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 538
QY 1026 tgtcgagcct 1035
Db 539 TGTCGCGCCT 548

RESULT 3
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
```

```
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 548 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..405
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1..405
; PCT-US95-07289-6

Query Match 42.2%; Score 515.6; DB 6; Length 548;
Best Local Similarity 98.2%; Pred. No. 3e-142;
Matches 540; Conservative 1; Mismatches 7; Indels 2; Gaps 2;

QY 486 atgagtcctgtgaaacaaatgtggcagagcctaaacatcgccctgggtgaatggaacc 545
Db 1 ATGAGTCCTGTGAAACAAATGTGGCAGAGCCCTAAACATCGCCCTGGTGAATGGAACC 60
QY 546 acgggagctgtgctgggacagaagcatttgacatgtactctgagatgttatgcaccta 605
Db 61 ACGGAGCTGTGCTGGACAGAGGCAATTTGACATGTACTCTGAGATGTTATGCACCTA 120
QY 606 gtgaattccctaaagaattccggggggtgacctgtgtgtgctgacctctacacgat 665
Db 121 GTGAAATTCCTTAAAGAAATTCGGGGGGTGCACCTGGTGGCTCCTACGACGAT 180
QY 666 ccagggaacaaatgaacgatgaagcagaaactcttctgagctgagagcccgagctctac 725
Db 181 CCAGGGACCAAAATGAACGATGAAGAGGAGGAACTCTTCTGACTTGGGGAGTTCTCTAC 240
QY 726 gcaaaacaactgggctccgggacagctgggtctctcatagagcgaacacccctcagggt 785
Db 241 GCAAAACAACCTGGGCTTCGGGGACAGCTGGGTCTTCATAGGAGCCCAAGACCTCAGGGT 300
QY 786 aaaagcccttgagcagctcttaaaagacagccagacacacacacacacacacacacac 845
Db 301 AAAAGCCCTTTGAGCAGTTCTTAAAGAACAGCCACAGACACAAACAAATACGAGGATGG 360
QY 846 ccagagctgtgagatgagctgacatgacatgacatgacatgacatgacatgacatgacat 905
Db 361 CCAGAGCTGTGAGATGAGGAGGCTGCATGCCGCCAGGCAATTTAGGGTGGCTGTGGC 420
QY 906 tcttctcagccaggggctgaaagagctgctgctgctgctgctgctgctgctgctgctgctg 965
Db 421 TCTTCTCCTCAGCCAGGGGCTGAAAGAGYTCCTGCTGCTCATTAGGAGTCANAGCCCGCA - 480
QY 966 gggctgagagagagcaggggtgctgctgagagagagagagagagagagagagagagagag 1025
Db 480 GGCTGNAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 538
QY 1026 tgtcgagcct 1035
Db 539 TGTCGCGCCT 548

RESULT 2
PCT-US95-07289-6
; General 6, Application PC/TUS9507289
; GENERAL INFORMATION:
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Rosen, Craig
; TITLE OF INVENTION: Colon Specific Genes and Proteins
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
; ADDRESSEE: Stewart & Olslein
; STREET: 6 Becker Farm Road
; CITY: Roseland
; STATE: NJ
; COUNTRY: USA
; ZIP: 07068-1739
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/07289
; FILING DATE: 06-JUN-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferraro, Gregory D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-265
```

```
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232,463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: pTZgpt-F1s
;
US-08-232-463-14
```

```
Query Match 3.8%; Score 46.6; DB 1; Length 7218;
Best Local Similarity 6.0%; Pred. No. 0.0006;
Matches 25; Conservative 213; Mismatches 177; Indels 0; Gaps 0;

Qy 480 atgatacgtgctgtgaaacaatgtggcagagcctaaacatcgccctgggtgaat 539
| || | : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1447 AAGAATTGGTACRrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr 1388

Qy 540 ggaaccaggagctgtgctgggacagaagcattgacatgtactctggagatgtatg 599
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1387 Rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr 1328

Qy 600 cacctagtgaattccttaaagaattccgggggtgactgtgtgtgtggtgctctac 659
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1327 Rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr 1268

Qy 660 gacgatccagggaccataatgaacagtgaagcaggaaactcttctgactggggagt 719
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1267 Rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr 1208

Qy 720 tcttacgcaaaactgggtctccgggacagctgggtcttctataggacaaagacctc 779
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1207 Rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr 1148

Qy 780 aggggtaaagcccttttgacagtctctaaagaacccagacacaaacaaatcacg 839
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1147 Rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr 1088

Qy 840 ggatggccagctgctggagatggaggtgcatgcccgcgaagccattttagg 894
: : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1087 Rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr 1033
```

```
RESULT 4
US-08-665-259-19
; Sequence 19, Application US/08665259
; Patent No. 6028173
; GENERAL INFORMATION:
; APPLICANT: Landes, Gregory M.
; APPLICANT: Burn, Timothy C.
; APPLICANT: Connors, Timothy D.
; APPLICANT: Dackowski, William R.
; APPLICANT: Van Raay, Terence J.
; APPLICANT: Klinger, Katherine W.
; TITLE OF INVENTION: NOVEL HUMAN CHROMOSOME 16 GENES,
; TITLE OF INVENTION: COMPOSITIONS, METHODS OF MAKING AND USING SAME
; NUMBER OF SEQUENCES: 73
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: One Mountain Road
; CITY: Framingham
; STATE: Massachusetts
; COUNTRY: United States of America
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/665,259
; FILING DATE: 17-JUN-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dugan, Deborah A.
; REGISTRATION NUMBER: 37,315
; REFERENCE/DOCKET NUMBER: IGS-9.1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (508) 872-8400
; TELEFAX: (508) 872-5415
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6803 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
;
US-08-665-259-19
```

```
Query Match 3.1%; Score 38; DB 5; Length 6803;
Best Local Similarity 59.4%; Pred. No. 0.19;
Matches 82; Conservative 0; Mismatches 55; Indels 1; Gaps 1;

Qy 868 gtgcatgcccccgaaagccattttagggtggctgtgtgtcttctcctcagccaggggcctga 927
||| |||| | : : : : : : : : : : : : : : : : : : : : : : : :
Db 158 GCTCCCTGCGCCTAGACTGCCAGCCCATCTATCCCTCTCCCGAGCCTGTGCCCAA 217

Qy 928 agaagctctgctgacttagagtcagagccggcaggggctgagggagagagcagg 987
||| |||| | : : : : : : : : : : : : : : : : : : : : : : : :
Db 218 AGCTGAGAGTCCACATC-TAGGGGTGAGGGGTGGGGTGGGGAGGGAGCGAAGCACATG 276

Qy 988 ggtgctgcatggaagtq 1005
||| |||| | : : : : : : : : : : : : : : : : : : : : : : : :
Db 277 CGCCTGAGTTGCAGGTG 294

RESULT 5
US-07-960-389-1
; Sequence 1, Application US/07960389
; Patent No. 5705611
; GENERAL INFORMATION:
; APPLICANT: HAYASHIDA, Kasuhiro;
; TITLE OF INVENTION: Human GM-CSF Receptor Component
; NUMBER OF SEQUENCES: 2
```



```

; APPLICANT: Piddington, Christopher S.
; APPLICANT: Rambosek, John A.
; APPLICANT: Turunen, Maria K.
; APPLICANT: Fagerstr m, Richard B.
; APPLICANT: Houston, Christine S.
; TITLE OF INVENTION: Production of Phytase Degrading Enzymes
; TITLE OF INVENTION: in Trichoderma
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1100 New York Avenue, Suite 600
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/609,426A
; FILING DATE: 01-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/923,724
; FILING DATE: 31-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/496,155
; FILING DATE: 19-MAR-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/044,077
; FILING DATE: 29-APR-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: UK 8610600
; FILING DATE: 30-APR-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Reed, Grant E.
; REGISTRATION NUMBER: P-41,264
; REFERENCE/DOCKET NUMBER: 1050.0080001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2540
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2363 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: both
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(404..447, 550..1906)
; US-08-609-426A-7

```

```

Query Match      2.8%; Score 34.4; DB 3; Length 2363;
Best Local Similarity 48.0%; Pred. No. 1.3;
Matches 98; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 252 gcctcatcttggccatagtcacgacatggtttatttcgaagtcacatgagcttcagc 311
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 989 GGCTCCAGCGCGGTGATCGCTCCGGGAGAAATTCATTGAGGCGTCCAGAGCACCAAG 1048
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 312 atgaataaccatcgctgtgccacgctggctgcctcccccacacgaagagatccagggttaa 371
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1049 CTGAAGGATCCTCGTGCCCGCCGCGGCAATCGTCGCCCAAGATCGACGTGGTCAATTCC 1108
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 372 aagtacaagtgtggcctcatcaagccctgccagccacactacttttcggtttaaatctgc 431
    || || || || || || || || || || || || || || || || || || || || ||
DB 1109 GAGGCCAGCTCATCCACAACTCTCGACCCGACCCGACCTGTCTGTTTGAAGACAGC 1168
    || || || || || || || || || || || || || || || || || || || || ||

QY 432 agtggggggcccaacgctgtgggc 455
    ||||| ||||| ||||| |||||

```

```

Db 1169 GAATTGGCGGATACCGTCGAAGCC 1192
RESULT 8
US-08-374-652C-1
; Sequence 1, Application US/08374652C
; Patent No. 5834286
; GENERAL INFORMATION:
; APPLICANT: NEVALAINEN, HELENA K.M.
; APPLICANT: PALOHEIMO, MARJA T.
; APPLICANT: FAGERSTROM, RICHARD B.
; APPLICANT: MIETTINEN-OINONEN, ARJA S.
; APPLICANT: TURUNEN, MARJA K.
; APPLICANT: RAMBOSEK, JOHN A.
; APPLICANT: PIDDINGTON, CHRISTOPHER S.
; APPLICANT: HOUSTON, CHRISTINE S.
; APPLICANT: CANTRELL, MICHAEL A.
; TITLE OF INVENTION: RECOMBINANT CELLS, DNA CONSTRUCTS,
; TITLE OF INVENTION: VECTORS AND METHODS FOR EXPRESSING PHYTASE DEGRADING
; TITLE OF INVENTION: ENZYMES IN DESIRED RATIOS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 NEW YORK AVENUE, SUITE 600
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/374,652C
; FILING DATE: 24-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07058
; FILING DATE: 27-JUL-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/925,401
; FILING DATE: 31-JUL-1992
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REED, GRANT E.
; REGISTRATION NUMBER: 41,264
; REFERENCE/DOCKET NUMBER: 1050.071001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2379 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(404..447, 550..1906)
; US-08-374-652C-1

```

```

Query Match      2.8%; Score 34.4; DB 3; Length 2379;
Best Local Similarity 48.0%; Pred. No. 1.3;
Matches 98; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 252 gcctcatcttggccatagtcacgacatggtttatttcgaagtcacatgagcttcagc 311
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1005 GGCTCCAGCGCGGTGATCGCTCCGGGAGAAATTCATTGAGGCGTCCAGAGCACCAAG 1064
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


[illegible]

RESULT 13

US-08-961-716-8/C
; Sequence 8, Application US/08961716
; Patent No. 5880272
; GENERAL INFORMATION:
; APPLICANT: ADACHI, HIDEKI
; APPLICANT: TSUJIMOTO, MASAFUMI
; APPLICANT: INOUE, KEIZO
; APPLICANT: ARAI, HIROYUKI
; TITLE OF INVENTION: OXIDIZED PHOPHOLIPID DEGRADING ENZYME
; TITLE OF INVENTION: AND GENE THEREOF
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER &
; ADDRESSEE: NEUSTADT P.C.
; STREET: 1755 S. Jefferson Davis Highway, Suite 400
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,716

```

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA: US 08/283,917
FILING DATE: 03-AUG-1994
APPLICATION NUMBER: JP 209943/1993
FILING DATE: 03-AUG-1993
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 5980272man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 2292-030-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 2085 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Bos taurus
FEATURE:
NAME/KEY: CDS
LOCATION: 844..2073

US-08-961-716-8

Query Match	2.8%;	Score 33.6;	DB 3;	Length 2085;
Best Local Similarity	56.2%;	Pred. No. 2.1;		
Matches 63;	Conservative	0;	Mismatches 49;	Indels 0;
				Gaps 0;

[illegible]

D_b 52 GGCCTGTCCCCCCTCCCTCCC GTGTC TCACTCAACGGCGT CGGCGGC 1

RESULT 12

US-08-151-574-33
; Sequence 33, Application US/08151574
; Patent No. 5436156
; GENERAL INFORMATION:
; APPLICANT: Robert F.M. Van Gorcom
; APPLICANT: Willem Van Hartingsveldt
; APPLICANT: Petrus A. Van Paridon
; APPLICANT: Annemarie E. Veenstra
; APPLICANT: Rudolf G.M. Luiten
; APPLICANT: Gerardus Luten
; TITLE OF INVENTION: Cloning and Expression of Microbial
; TITLE OF INVENTION: Phytase
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESS: Morrison & Foerster
; STREET: 545 Middlefield Road, Suite 200
; CITY: Menlo Park
; STATE: California
; COUNTRY: USA
; ZIP: 94025-3471
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; SOFTWARE:
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08151,574
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/688,578
; FILING DATE: 24-MAY-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H.
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 24615-20026.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-327-7250
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1404 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Aspergillus ficum (Aspergillus niger)
; STRAIN: NRRL 3135
; US-08-151-574-33

Query Match	2.7%;	Score 33;	DB 1;	Length 1404;
Best Local Similarity	45.3%;	Pred. No. 2.6;		
Matches 120:	Conservative	0;	Mismatches 145;	Indels 0;
			Gaps 0;	

Qy	252	gcctcatcttggccatagtcacagacatgatgtttatttgaagctacatcagcttcacg	311
Db	484	GGCTCCAGCCGGTGATCGCTCCGGCAAGAAATTCATCAGGGCTTCCAGAGCACCAAG	543
Qy	312	atgaaaaaccatccgtctgccaagctggtcgtgcgtcccaagagagatccaaggttaa	371
Db	544	CTGAAGGATCCTCGTGGCCCGCGGCAATCGTCGCCCAAGATCGAGCTGGTCATTTC	603
Qy	372	aagtacaagtggtgcctcatcaagcctgcccagccaactacittgcgtttaaactcgc	431
Db	604	GAGGCCAGCTCATCCAACAACACTCTCGACCCAGGACCTGCACTGTCCTCGAAGACAG	663

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 31, 2000, 16:54:03 ; Search time 799.5 Seconds
(without alignments)
6190.101 Million cell updates/sec

Title: US-09-186-342-1
Perfect score: 1221
Sequence: 1 cgcccgggcaggtgcattct.....aaattttatttttgcggta 1221

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues 9714632
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

- EST:*
- 1: em_est1:*
 - 2: em_est2:*
 - 3: em_est3:*
 - 4: em_est4:*
 - 5: em_est5:*
 - 6: em_est6:*
 - 7: em_est7:*
 - 8: em_est8:*
 - 9: em_est9:*
 - 10: em_est10:*
 - 11: em_est11:*
 - 12: em_est12:*
 - 13: em_est13:*
 - 14: em_est14:*
 - 15: em_est15:*
 - 16: em_est16:*
 - 17: em_est17:*
 - 18: em_est18:*
 - 19: em_est19:*
 - 20: gb_est1:*
 - 21: gb_est2:*
 - 22: gb_est3:*
 - 23: gb_est4:*
 - 24: gb_est5:*
 - 25: gb_est6:*
 - 26: gb_est7:*
 - 27: gb_est8:*
 - 28: gb_est9:*
 - 29: gb_est10:*
 - 30: gb_est11:*
 - 31: gb_est12:*
 - 32: gb_est13:*
 - 33: gb_est14:*
 - 34: gb_est15:*
 - 35: gb_est16:*
 - 36: gb_est17:*
 - 37: gb_est18:*
 - 38: gb_est19:*
 - 39: gb_est20:*
 - 40: gb_est21:*
 - 41: gb_est22:*
 - 42: gb_est23:*
 - 43: gb_est24:*
 - 44: gb_est25:*

- 45: gb_est26:*
- 46: gb_est27:*
- 47: gb_est28:*
- 48: gb_est29:*
- 49: gb_est30:*
- 50: gb_est31:*
- 51: gb_est32:*
- 52: em_est20:*
- 53: em_est21:*
- 54: em_est22:*
- 55: em_est23:*
- 56: em_est24:*
- 57: em_est25:*
- 58: em_est26:*
- 59: gb_est33:*
- 60: gb_est34:*
- 61: gb_est35:*
- 62: gb_est36:*
- 63: gb_est37:*
- 64: gb_est38:*
- 65: em_est27:*
- 66: em_est28:*
- 67: em_est29:*
- 68: em_est30:*
- 69: gb_est39:*
- 70: gb_est40:*
- 71: gb_est41:*
- 72: gb_est42:*
- 73: gb_est43:*
- 74: gb_est44:*
- 75: em_est31:*
- 76: em_est32:*
- 77: em_est33:*
- 78: em_est34:*
- 79: gb_est45:*
- 80: gb_est46:*
- 81: gb_est47:*
- 82: gb_gss1:*
- 83: gb_gss2:*
- 84: gb_gss3:*
- 85: gb_gss4:*
- 86: em_gss1:*
- 87: em_gss2:*
- 88: em_gss3:*
- 89: em_gss4:*
- 90: gb_gss5:*
- 91: gb_gss6:*
- 92: gb_gss7:*
- 93: gb_gss8:*
- 94: gb_gss9:*
- 95: em_gss5:*
- 96: em_gss6:*
- 97: em_gss7:*
- 98: em_gss8:*
- 99: em_gss9:*
- 100: em_gss10:*
- 101: em_gss11:*
- 102: gb_gss10:*
- 103: gb_gss11:*
- 104: em_gss12:*
- 105: gb_gss12:*
- 106: gb_gss13:*
- 107: gb_gss14:*
- 108: gb_gss15:*
- 109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result % Query

No.	Score	Match	Length	DB	ID	Description
C 1	624.8	51.2	645	79	AW300770	AW300770 xk05q02.x
C 2	616	50.5	668	63	AW001287	AW001287 wf27e06.x
C 3	595.6	48.8	615	50	AI691072	AI691072 wf22c03.x
C 4	568.4	46.6	646	62	AI936111	AI936111 wo61h03.x
C 5	558	44.1	564	36	AA622758	AA622758 np76c04.s
C 6	522.2	42.8	538	44	AI245950	AI245950 qk45b04.s
C 7	518	42.4	518	35	AA563933	AA563933 nk23b02.s
C 8	498	40.8	498	45	AI348065	AI348065 qp56d09.x
C 9	496.8	40.7	594	35	AA552519	AA552519 nk14e01.s
C 10	495.4	40.6	508	60	AI801582	AI801582 tg91c11.x
C 11	495	40.5	518	36	AA622120	AA622120 ng56b03.s
C 12	471.6	38.6	479	63	AW001308	AW001308 wu28c10.x
C 13	464	38.0	464	39	AA847242	AA847242 of01b11.s
C 14	456.2	37.4	461	36	AA622570	AA622570 np21f04.s
C 15	449.4	36.8	454	35	AA552362	AA552362 nk04h09.s
C 16	449	36.8	455	49	AI660557	AI660557 we68b05.x
C 17	424	34.7	530	64	AW050790	AW050790 wz21e07.x
C 18	402	32.9	417	35	AA582787	AA582787 nm35b01.s
C 19	401.8	32.9	443	63	AW000826	AW000826 wu45d09.x
C 20	371.4	30.4	396	36	AA643708	AA643708 nl96f08.s
C 21	370	30.3	375	51	AI732367	AI732367 nf61e11.x
C 22	356.4	29.2	434	39	AA857546	AA857546 of64c07.s
C 23	355	29.1	360	34	AA514424	AA514424 nf61e11.s
C 24	355	29.1	364	36	AA643616	AA643616 ng79b08.s
C 25	352	28.8	358	50	AI673534	AI673534 we75b10.x
C 26	320.8	26.3	325	35	AA543029	AA543029 nf92e04.s
C 27	310.4	25.4	321	61	AI821215	AI821215 ne12h10.y
C 28	307	25.1	337	27	AA025434	AA025434 ze84f10.s
C 29	302	24.7	306	31	AA297147	AA297147 EST112731
C 30	286	23.4	301	34	AA470683	AA470683 ne12h10.s
C 31	283.8	23.2	302	51	AI732198	AI732198 ne12h10.x
C 32	274.4	22.5	289	48	AI582013	AI582013 ar98b03.x
C 33	267	21.9	268	31	AA297176	AA297176 EST112740
C 34	256.2	21.0	417	31	AA298484	AA298484 EST114112
C 35	246	20.1	249	51	AI749731	AI749731 at32a08.x
C 36	230	18.8	461	36	AA611725	AA611725 vo89f11.r
C 37	223	18.3	379	80	AA315283	AA315283 12472 MAR
C 38	216.2	17.7	377	27	AA025433	AA025433 ze84f10.r
C 39	209.4	17.1	387	33	AA451058	AA451058 vf86b07.r
C 40	181	14.8	778	62	AI883327	AI883327 fc60e01.y
C 41	180.6	14.8	525	80	AA323823	AA323823 uo77a04.y
C 42	176.8	14.5	264	38	AA793766	AA793766 vf96e04.r
C 43	160.8	13.2	311	80	AA315346	AA315346 12597 MAR
C 44	151.6	12.4	410	24	N41338	N41338 yw68q05.r1
C 45	151.4	12.4	509	34	AA512284	AA512284 vj19g03.r

ALIGNMENTS

```

RESULT 1
AW300770/c 645 bp mRNA EST 18-JAN-2000
LOCUS xk05q02.x1 NCI CGAP_Col9 Homo sapiens CDNA clone IMAGE:2665874 3'
DEFINITION similar to SM:EF07_MOUSE F97805 PROTEIN EF-7 ;, mRNA sequence.
ACCESSION AW300770.1 GI:6710447
VERSION EST.
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 645)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Jul 8, 1999 this sequence version replaced gi:5420873.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov

```

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center

Clone Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/brrp/image/image.html

Possible reversed clone: polyT not found
Seq primer: -40UP from Gibco
High quality sequence stop: 403.

FEATURES

source
Location/Qualifiers
1..645
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2665874"
/clone_lib="NCI-CGAP_Col9"
/tissue_type="moderately differentiated adenocarcinoma"
/lab_host="DH10B"
/notes="Organ: colon; Vector: pCMV-SPORT6; Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Normalized to Cot 50. Average insert size 1.32kb. Normalized version of NCI-CGAP_Col8. Library constructed by Life Technologies."
BASE COUNT 137 a 192 c 159 g 156 t 1 others
ORIGIN

Query Match 51.2%; Score 624.8; DB 79; Length 645;
Best Local Similarity 98.0%; Pred. No. 3.9e-160;
Matches 632; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

```

QY 576 gacatgactctgagatgttatgcacactagtgaataatctctaaagaatactcgggggggt 635
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 645 GACATGATTCTGAAGATGTATGCACNTAGTGAATTTCTTAAAGAAATTCGCGGGGT 586

QY 636 gcaatggtgctgggtgctctacgacgatccaggggaccacaaatgaacgatgaacgag 695
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 585 GCACGTGCTGTGGTGGCTCTTACGCCGATCCAGGGACCAAAATGAACGATGAACGAGA 536

QY 596 aaactctctctgactctgagggtctctacgaaacaactcggggtccggggacagtgg 755
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 525 AAATCTTCTGACTTTGGGAGTTCTTACGCAAAAACAACTGGGCTTCCGGGACAGCTGG 466

QY 756 gtcttcataagaccagacacctcagggttaaaagccctttgagcagttctctaaagaac 815
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 465 GTTTCATAGGAGCCAAAGACCTCAGGGGTAAAGCCCTTTGAGCAGTTCCTTAAAGAAC 406

QY 816 agccagacacacacaaatacaggggagtgccagagctgctggagatgagggctgcatg 875
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 405 AGCCAGACACAAACAAATACGAGGGATGCCAGAGCTGCTGGAGATGGAGGCTGCATG 346

QY 876 ccccggaagccattttagggtggtgctgtctctcctcagccagggggtcctgaagaagctc 935
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 345 CCCCAGAGCCATTTTAGGTGGCTGTGGCTCTTCTCAGCCAGGGGCTGAAAGAGCTC 286

QY 936 ctgctgacttagagtcagagcccgaggggctctagagagagagcagggggtgctac 995
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 285 CTGCCTGACTTAGAGTCAGAGCCCGGCGGCTGAGGAGGAGGAGCGGGGTCTGCTGC 226

QY 996 gtggaaggtgctcaggtccttgacgctgtgtcgcgctctctcctcctcctcggaacagaa 1055
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 225 GTGAAGGTGCTGCAGGTCTCTTGCACGCTGTGTCGCGCTCTCTCTCTCGGAACAGAAC 166

QY 1056 cctccacagacacatctaccgccgaagacagcctcagagggtcctcttggaaccagctg 1115
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 165 CCTCCACACACATCTTACCCGGAAGACAGCCTCAGAGGGTCTCTTCTGGAACACAGCTG 106

QY 1116 tctgtgagagaatgggtgcttcttcgtcagggactgctgacggtggtcctcctgaggaaga 1175
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 105 TCTGTGAGAGATGGGGTCTTTCTGTCAGGGACTCTGCTGACGGCTGCTCTGAGGAGGA 46

```

QY 1176 caaactgccagacttgagcccaattaaattttattttgtcgtggt 1220
|||||
Db 45 CAAACTGCCAGACTTGAGCCCAATTAAATTTTATTTTTGTGTTG 1

RESULT	2	
LOCUS	AW001287/c	
DEFINITION	668 bp mRNA EST 27-OCT-1999 wu27e08.x1 Soares_bieckgraeft_colon_NHCD Homo sapiens cDNA clone IMAGE:2521282 3', similar to SW:BF07 MOUSE P97805 PROTEIN EF-7 :contains MSR1.t2 MER22 repetitive element ;, mRNA sequence.	
ACCESSION	AW001287	
VERSION	AW001287.1	GI:5848203
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 668)	
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .	
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP).	
JOURNAL	Tumor Gene Index	
COMMENT	Unpublished (1997) On Mar 9, 1996 this sequence version replaced gi:1132811.	

Query Match	50.5%	Score 616;	DB 63;	Length 668;
Best Local Similarity	96.8%;	Pred. No. 1e-157;		
Matches 639:	Conservative	0: Mismatches	20: Indels	1: Gaps

Qy	562	gacagaagcattgacatgtactctctggagatgttatgcacctagtcgaattctcttaag	621
Db	667	gccaccagcatttgacatgcacttaggagaagtattgcacacctagtgaaattccataaga	608
Qy	622	aaattccgggggtgcactgctgctgtgctcctctacgacgtatccaggagaccataatga	581
Db	607	tatttccgggggtgctgctgctgtgtgtgtggcctctacgacgattccaggaccataatga	548
Qy	682	acgatgaagcagaaactctctcactctggggagcttctacgcgaatacaactgggct	741
Db	547	tgcattgaagcagaaactctctcactctggggagcttctacgcgaatacaactgggct	488


```

Qy 1032 gctctctctcgtgaagaacacccctccacagacacatctaccctggaagaccagctc 1091
Db 158 GCCTCTCTCTCGTGAAGAACACCCCTCCACACACATCTACCCGGAAGACCGCTC 99
Qy 1092 agagggctctctggaaccagctctctgtgtgagagaatgggtgtcttctcaggagcty 1151
Db 98 AGAGGGCTCTTCTGTAACACAGCTCTCTGTGTGAGAGAATGGGTCTTCTGTCAGGACTG 39
Qy 1152 ctgacgctgtctctggaagagacaaactgccagac 1189
Db 38 CTGACGGCTGTCTGTGAGGAGGACAAACTGCCCAGAC 1

RESULT 8
AI348065/c
LOCUS AI348065 498 bp mRNA EST 02-FEB-1999
DEFINITION qp56d09.x1 NCI_CGAP_Co8 Homo sapiens cDNA clone IMAGE:1927025 3',
similar to TR:P97805 P97805 EF-7 ; mRNA sequence.
ACCESSION AI348065
VERSION AI348065.1 GI:4085271
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 498)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Feb 6, 1998 this sequence version replaced gi:2848226.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 841 Std Error: 0.00
Seq primer: -40UP from Gibco
High quality sequence stop: 459.
Location/Qualifiers
1. .498
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1927025"
/clone_lib="NCI_CGAP_Co8"
/tissue_type="adenocarcinoma"
/lab_host="DH10B"
/notes="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
modified polylinker; 1st strand cDNA was prepared from
colon adenocarcinoma, and was then primed with a Not I -
oligo(dT) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT7T3
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."
BASE COUNT 98 a 154 c 131 g 115 t
ORIGIN

Query Match 40.8%; Score 498; DB 45; Length 498;
Best Local Similarity 100.0%; Pred. No. 1.5e-125;
Matches 498; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 723 tacgcaaaacactgggtctccggagcagctgggtctctcattagagcgaagacctcagg 782
Db 498 TACGCAAAACACTGGGCTCCGGAGCAGCTGGGTCTCTATAGGAGCAAGACCTCAGG 439

```

```

Qy 783 ggtaaaagccctttgagcagttctttaagaagacccagacacacaaatacagagga 842
Db 438 GGTAAAAGCCCTTTGAGCAGTCTTAAAGAACAGCCAGACACAAATAACGAGGA 379
Qy 843 tggccagagctgctgagatggagggctcatgccccgaagccatttttaggtgctgt 902
Db 378 TGGCCAGAGCTGCTGGAGATGGAGGGCTCATGCCCGAAGCCCATTTAGGGTGGCTGT 319
Qy 903 ggtcttctcagccaggggctggaagctcctgctgacttaggtgagtcagcccg 962
Db 318 GGCTCTTCTCAGCCAGGGGCTCTGAAGAAAGCTCCTCCCTGACTTAGGAGTCAGAGCCGG 259
Qy 963 caggggctgagagagagcaggggtgtctgtgtggaaggtgtcaggtcctctgacg 1022
Db 258 CAGGGGCTGAGGAGGAGGAGGAGGGGTGCTGCTGTGAAGGTGCTGCAGGTCTTGTGCT 199
Qy 1023 ctgtgtcgcgcctctctctcctcgaaacagacccctccacagcacatctaccctcgaag 1082
Db 198 CTGTGTGCGGCTCTCTCTCTCGGAAACAGAACCCCTCCACAGACACATCTTACCCGGAAG 139
Qy 1083 accagctcagaggggtcctcttggaaccagctgtctgtgtgagagaatgggtgtcttctgt 1142
Db 138 ACCAGCTCAGAGGGTCTCTTCTGGAACCAAGCTGTCTGTGGAGAGAATGGGTGCTTGTCT 79
Qy 1143 caggactctcagcgtgtgtctcctcaggaagagacaaactgccagacttgaagcccaatta 1202
Db 78 CAGGGACTGCTGACGCTGTGTCTGTGAGGAAGGACAAACTGCCAGACTTGAGCCCAATTA 19
Qy 1203 aattttatttttctgtgt 1220
Db 18 AATTTATTATTTTCTGTGT 1

RESULT 9
AA552519/c
LOCUS AA552519 594 bp mRNA EST 05-SEP-1997
DEFINITION nk14e01.s1 NCI_CGAP_Co2 Homo sapiens cDNA clone IMAGE:1013456 3',
similar to TR:G452270 G452270 2-19 PROTEIN PRECURSOR. ; mRNA
sequence.
ACCESSION AA552519
VERSION AA552519.1 GI:2322773
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 594)
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT On Nov 29, 1993 this sequence version replaced gi:502047.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.
Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: Stratagene, Inc., David B. Krizman,
Ph.D.
cDNA Library Arraying: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1404 Std Error: 0.00
Seq primer: -40ml3 fwd. Et from Amersham
High quality sequence stop: 347.
Location/Qualifiers
1. .594
/organism="Homo sapiens"
/db_xref="taxon:9606"

```


Db	29	TTGAGCCCAATTAAATTTATTTTCTG	1
RESULT	11		
AA622120/c			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
BASE COUNT			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
QY	715	ggagttcctacgc-aaacaactgggtctccggagcagctgggtcttcagagccaaa	773
Db	518	GGAGTTCTCCTACGAAAAAACCTGGGCTTCGGGACAGCTGGGTCTCTATAGAGGCCAAA	459
QY	774	gacctcgggtgtaaaagccctttgagcagttcttaagaacagccagcacacaacaaa	833
Db	458	GACCTCAGGGGTAAGACCCCTTTGACAGTCTTAAAGACAGCCAGCACACAAACAA	399
QY	834	tacgagggatggccagagctgctggagatggaggggtgcgtccccccgaagccatttag	893
Db	398	TACGAGGGATGGCCAGAGCTGCTGGAGATGGAGGTGCGATGCCGCCGAAGCCATTTAG	339

BASE COUNT	93 a	149 c	125 g	112 t								
ORIGIN												
Query Match	38.6%;	Score 471.6;	DB 63;	Length 479;								
Best Local Similarity	99.2%;	Pred. No. 2.4e-118;										
Matches 474;	Conservative 0;	Mismatches 4;	Indels 0;	Gaps 0;								
QY	730 acaactgggtcccgagcagctgggtctctcattagagcacaagacccctcaggggtataaa 789											
Db	478 AACAACTGGGCTTCGGGACAGCTGGTCTCTATAGAGGCCAAAGACCTCAGGGGTAAA 419											
QY	790 gccctttgagcagtgctttaaagaacagccagacacacaacaataacagggatggccag 849											
Db	418 GCCCCTTTGAGCAGTTCTTAAAGAACAGCCAGACACAAACAAATACGAGGATGCCAG 359											
QY	850 agctgctggagatggagggtcgtcatgccccgaagcatttttaggtggtgctgctctt 909											
Db	358 AGCTGCTGGAGATGGAGGGCTGATGCCCAAGGCCATTTTAGGGTGGCTGTGGCTTTT 299											
QY	910 cctcagccagggcctgaagaagctcctgctgacttagagtcagagcccgaggggc 969											
Db	298 CCTCAGCCAGGGCCTTAAGAAGTTCTGCTGACTTAGAGTACAGAGCCCGCAGGGGC 239											
QY	970 ttagggaggaagcaggggtgctgctggaaggtgctgcaggtccttgcaagctgtgtc 1029											
Db	238 TGAGGAGGAGGAGCAGGGGTGCTGCTGGAAGGTGCTGCAGGTCTTGCACGCTGTGTC 179											
QY	1030 gcgcctctctcgtgaaacagacccctccacagacacatcctaccgggaagaccagcc 1089											
Db	178 GCCTCTCTCTCGGAACAGAACCCCTCCACAGCATCTCTACCCGGAAGACAGCC 119											
QY	1090 tcagaggggtccttgaaccagctgtctgtggagagaatgggtgcttctcaggggac 1149											
Db	118 TCAGAGGGTCTTCTGGAACACAGCTCTGTGTGGAGAGATGGGGTCTTCTCAGGGAC 59											
QY	1150 tgcctgagcgtgctcctgagggaagacaactgcccagacttgagcccaatttaattt 1207											
Db	58 TGTGAGCGGTGCTGCTGTGAGGAAGGACAAACTGCCAGACTTGTAGCCCAATTAATTT 1											
RESULT 13												
AA847242/c												
LOCUS	AA847242	464 bp	mRNA	EST	31-MAR-1998							
DEFINITION	of01b11.s1 NCI_CGAP_Col2 Homo sapiens cDNA clone IMAGE:1419837 similar to TR:P97805 P97805 EF-7 ;, mRNA sequence.											
ACCESSION	AA847242											
VERSION	AA847242.1	GI:2933760										
KEYWORDS	EST.											
SOURCE	human.											
ORGANISM	Homo sapiens											
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.											
AUTHORS	1 (bases 1 to 464)											
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index											
JOURNAL	Unpublished (1997)											
COMMENT	On Jan 19, 1998 this sequence version replaced gi:2153278. Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1350 Email: Robert.Strausberg@nih.gov Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Stratagene, Inc. DNA Sequencing by: Washington University Genome Sequencing Center Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: www-bio.llnl.gov/bbrp/image/image.html											
Insert Length:	1265	Std Error:	0.00									
Seq primer:	-40m13 fwd.	ET from	Amersham									
FEATURES												
Source	High quality sequence stop: 453. Location/Qualifiers											
	1..464											
	/organism="Homo sapiens"											
	/db_xref="taxon:9606"											
	/clone_image="IMAGE:1419837"											
	/clone_lib="NCI_CGAP_Col2"											
	/sex="mixed"											
	/tissue_type="colon tumor"											
	/lab_host="SOLR (kanamycin resistant)"											
	/note="organ: colon; Vector: Bluescript SK-; Site: 1: EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Pooled colon tumors. 5' adaptor sequence: 5' GAATTCGCGACGAG 3' 3' adaptor sequence: 5' CTCGAGTGTGTGTGTGTGTGT 3' Average insert size: 1.2 kb."											
BASE COUNT	82 a	149 c	126 g	107 t								
ORIGIN												
Query Match	38.0%;	Score 464;	DB 39;	Length 464;								
Best Local Similarity	100.0%;	Pred. No. 2.8e-116;										
Matches 464;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;								
QY	726 gcaaaacaactgggtcccgagcagctgggtctctcattagagcacaagacccctcaggggt 785											
Db	464 GCAAAACAACCTGGGCTTCGGGACAGCTGGTCTCTATAGAGGCCAAAGACCTCAGGGGT 405											
QY	786 aaagcccccttgagcagttcttaaaagacagccagacacaacaataacgagggatgg 845											
Db	404 AAAAGCCCTTTGAGCAGTTCTTAAAGAACAGCCAGACACAAACAAATACGAGGATGG 345											
QY	846 ccagagctgctgagatggaggctgcatgccccgaagcatttttaggtggtgtggc 905											
Db	344 CCAGAGCTGCTGGAGATGGAGGGCTGATGCCCCGAAAGCCATTTTAGGGTGGCTGTGGC 285											
QY	906 tcttctcagcaggggctgaaagagctcctgctgacttaggtcagagcccgagcag 965											
Db	284 TCTTCTCAGCAGGGGCTCGAAGAGCTCCTGCTGCTTAGGAGTCAGAGCCCGGCAG 225											
QY	966 gggctgaggaaggagcaggggggtgctgctggaaggtgctgaggtccttgcacgctg 1025											
Db	224 GGGCTGAGGAGGAGGAGCAGGGGTGCTGCTGGAAGGTGCTGCAGGTCTTGCACGCTG 165											
QY	1026 tgcgcgctctctcctcgtgaaacagacccctccacagacacatcctaccggaagacc 1085											
Db	164 TGTGCGGCTCTCTCTCTCGGAACACACACCTCCACAGACATCTCTACCCGGAAGACC 105											
QY	1086 agcctcagaggggtccttcttgaaaccagctgctgtggagagaatgggtgcttcttcgacg 1145											
Db	104 AGCCTCAGAGGGTCTTCTGGAACACAGCTGCTGTGTGGAGAGATGGGGTGTCTTCGTCAG 45											
QY	1146 ggaactgtgacggctgctcctgaggaaggacaactgccccagac 1189											
Db	44 GGACTGTGAGGGTGTCTCTGTGAGGAAGGACAAACTGCCAGAC 1											
RESULT 14												
AA622570/c												
LOCUS	AA622570	461 bp	mRNA	EST	21-OCT-1997							
DEFINITION	np211f04.s1 NCI_CGAP_GC5 Homo sapiens cDNA clone IMAGE:1116991 3' similar to TR:G452270 G452270 2-19 PROTEIN PRECURSOR. ;, mRNA sequence.											
ACCESSION	AA622570											
VERSION	AA622570.1	GI:2526446										
KEYWORDS	EST.											
SOURCE	human.											
ORGANISM	Homo sapiens											
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.											
AUTHORS	1 (bases 1 to 461)											
TITLE	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap. National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index											

JOURNAL
COMMENT

Unpublished (1997)
On Sep 29, 1997 this sequence version replaced gi:1520561.
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Stratagene, Inc.
DNA Sequencing by: Greg Lennon, Ph.D.
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 681 Std Error: 0.00
Seq primer: -40ml3 fwd. ET from Amersham
High quality sequence stop: 381.

FEATURES

source

1. 461
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:1116991"
/clone_lib="NCI_CGAP_GC5"
/tissue_type="germ cell tumor"
/lab_host="SOLR (kanamycin resistant)"
/note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2:
XhoI; Cloned unidirectionally. Primer: Oligo dt. Mixed
germ cell tumors. 5' adaptor sequence: 5' GAATTCGGCAGAG
3' 3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'
Average insert size: 0.7 kb."
84 a 146 c 125 g 106 t

BASE COUNT

ORIGIN

Query Match 37.4%; Score 456.2; DB 36; Length 461;
Best Local Similarity 99.3%; Pred. No. 3.8e-114;
Matches 458; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 729 aacaaactgggtcccgagctgggtttctataggagccaaagacctcagggttaa 788
DB 461 AAACAACCTGGCTTCCGTCAGCTGGCTTCTATAGGAGCCAAAGACCTTCAGGGGTAA 402

QY 789 agccctttgagcaggtcttaagaacagccagacacaacaatacagggaatgcc 848
DB 401 AGCCCCCTTGAGCAGTCTTAAAGAACAGCCCCCAGACACAACAATACGAGGGATGCCA 342

QY 849 gagctgctggagatggagggtgctgctgcccgaagccatttaggggtgctgtgctct 908
DB 341 GAGCTGCTGGAGATGGAGGGCTGCATGCCCCGGAAGCCATTTAGGGTGGCTGTGCTCT 282

QY 909 tctcagccagggcctgaagaagctcctgctgacttaggtcagagcccgagggg 968
DB 281 TCCTCAGCCAGGGCCCTGAAGAAGCTCTGCTGACTTAGGAGTCAGAGCCCGCAGGGG 222

QY 969 ctgaggagagacagaggggtgctgctggaagtactcaggtccttcacagctatgt 1028
DB 221 CTGAGGAGGAGGAGCAGGGGTGCTGCTGGAAGGTCTGCAGGTCTCTGCACGCTGTGT 162

QY 1029 cgcgctctctctcgaacacagaacctcccacagacacatctaccggaagaccagc 1088
DB 161 CGCGCTCTCTCTCGGAACACAGACCCCTCCACAGACATCTACCCGGAGACCAGC 102

QY 1089 ctgagaggtctcttctggaaccagctgtctgtgagagaatggggtcttctcaggg 1148
DB 101 CTCAGAGGGTCTCTGTGAACACAGCTGTCTGTGGAGAGATGGGGTCTTTTCGTACAGGA 42

QY 1149 ctgctgacgctgtgctctgagaaaggacaaactgcccagac 1189
DB 41 CTGCTGACGGTGTCTCTGAGGAAGGACAAACTGCCCAGAC 1

RESULT 15
AA52362/cLOCUS
DEFINITION

AA52362 454 bp mRNA EST 05-SEP-1997
nk04h09.s1 NCI_CGAP_Co2 Homo sapiens cDNA clone IMAGE:1012577 3'
similar to TR:G452270 G452270 2-19 PROTEIN PRECURSOR. ;, mRNA
sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert.Strausberg@nih.gov

Tissue Procurement: L. Jeffrey Medeiros, M.D., Michael R.

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: Stratagene, Inc., David B. Krizman,

Ph.D.

CDNA Library Arraying: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:

www-bio.llnl.gov/bbrp/image/image.html

Insert Length: 1355 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham

High quality sequence stop: 446.

Location/Qualifiers

1. 454

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:1012577"

/clone_lib="NCI_CGAP_Co2"

/tissue_type="tumor"

/lab_host="SOLR (kanamycin resistant)"

/note="Organ: colon; Vector: Bluescript SK-; Site_1:

EcoRI; Site_2: XhoI; Cloned unidirectionally. Primer:

Oligo dt. Bulk colon villous adenoma. 5' adaptor sequence:

5' GAATTCGGCAGAG 3' 3' adaptor sequence: 5'

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

CTCGAGTTTTTTTTTTTTTTT 3' Average insert size: 1.1 kb."

```

Qy 1068 catctaccocgggaagaccagcctcagaggggtccttcttggaaccagctgtctgtggagaga 1127
Db      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
154 CATCTACCCGGGAAGACCAGCCTCAGAGGGTCCTTCTGTGGAACCCAGCTGTCTGTGGAGAGA 95
Qy 1128 atgggggtccttcaggaactgctgacggctggctcctgaggaagacaaaactgcccag 1187
Db      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
94 ATGGGGTGCTTTCGTCAGGACTGCTGACGGCTGGTGCTCTGAGGAGGACAAACTGCCCCAG 35
Qy 1188 acttgagcccaattaaattttttttgtg 1218
Db      |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
34 ACTTGACCCCAATTAAATTTTATTGCTG 4

```

Search completed: May 31, 2000, 17:51:15
Job time: 3432 sec

WQSRFH (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 31 11:04:55 2000; Maspar time 21.50 Seconds
Tabular output not generated. 719.210 Million cell updates/sec

Title: >US-09-186-342-2
Description: (1-223) from US09186342.pep
Perfect Score: 1686
Sequence: 1 MRVSGVLRLLALIFAIVTW.....KYEGWPELLEMGCMPPKPF 223

Scoring table:
PAM 150
Gap 11

Searched: 225878 seqs, 69334122 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: sptrmb112
1:sp-archaea 2:sp-bacteria 3:sp-fungi 4:sp-human
5:sp-invertebrate 6:sp-mammal 7:sp-mhc 8:sp-organelle
9:sp-phase 10:sp-plant 11:sp-rodent 12:sp-unclassified
13:sp-vertebrate 14:sp-virus

Statistics: Mean 45.504; Variance 84.613; scale 0.538

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description	Pred. No.
1	110	6.5	44	11	HYPOTHETICAL 5.0 KD PR	2.54e-02
2	110	6.5	3190	5	CREB-BINDING PROTEIN H	2.54e-02
3	108	6.4	1001	5	TWO P DOMAIN POTASSIUM	4.90e-02
4	103	6.1	358	5	MOIE5.2 PROTEIN.	2.44e-01
5	101	6.0	253	1	HYPOTHETICAL 30.2 KD P	4.58e-01
6	102	6.0	2285	5	KAKAPO (FRAGMENT).	3.35e-01
7	102	6.0	2396	5	KAKAPO (FRAGMENT).	3.35e-01
8	100	5.9	510	5	INOSINE-5'-MONOPHOSPHA	6.25e-01
9	99	5.9	531	6	UDP-GLUCURONOSYLTRANSF	8.52e-01
10	99	5.9	532	6	UDP-GLUCURONOSYLTRANSF	8.52e-01
11	98	5.8	386	2	XYLOSE REPRESSOR.	1.16e+00
12	97	5.8	396	11	HSA BINDING PROTEIN 3.	1.57e+00
13	97	5.8	561	2	PHI SYNTHASE.	1.57e+00
14	97	5.8	1134	1	MOLYBDOTERIN OXIDORED	1.57e+00
15	96	5.7	269	5	COSMID C25H3.	2.12e+00
16	96	5.7	320	5	SIMILARITY TO FREQUENI	2.12e+00
17	96	5.7	389	10	PUTATIVE N-MYRISTOYLTR	2.12e+00
18	95	5.6	124	2	HYPOTHETICAL 14.5 KD P	2.87e+00
19	95	5.6	233	3	N-(5'-PHOSPHORIBOSYL)-	2.87e+00
20	94	5.6	240	2	INNER MEMBRANE PROTEIN	3.86e+00

21	95	5.6	350	5	Q17586	C01H6.6 PROTEIN.	2.87e+00
22	94	5.6	547	5	Q23135	SIMILAR TO GAMMA-GLUTA	3.86e+00
23	94	5.6	556	4	Q93000	CHL1 PROTEIN (FRAGMENT	3.86e+00
24	94	5.6	642	4	Q14777	RETINOBLASTOMA-ASSOCIA	3.86e+00
25	94	5.6	906	4	Q92770	HELICASE.	3.86e+00
26	95	5.6	906	4	Q92998	CHL1 POTENTIAL HELICAS	3.86e+00
27	95	5.6	1018	5	Q17874	F46F6.2 PROTEIN.	2.87e+00
28	92	5.5	178	2	Q48283	BACTERIOPHAGE-LIKE GEN	6.95e+00
29	92	5.5	231	5	Q16239	CL3D9.5 PROTEIN.	6.95e+00
30	92	5.5	339	5	Q45064	B0212.4 PROTEIN.	6.95e+00
31	92	5.5	350	5	Q19556	F18E2.4 PROTEIN.	6.95e+00
32	93	5.5	469	2	Q66887	HYPOTHETICAL 54.8 KD P	5.18e+00
33	92	5.5	509	3	Q06137	SIMILAR TO FRUCTOSE-2.	6.95e+00
34	93	5.5	533	6	Q46548	UDP-GLUCURONOSYLTRANSF	5.18e+00
35	93	5.5	533	6	Q46549	UDP-GLUCURONOSYLTRANSF	5.18e+00
36	93	5.5	573	2	Q34989	YVRG PROTEIN.	5.18e+00
37	92	5.5	805	5	Q17235	K10F12.5 PROTEIN.	6.95e+00
38	92	5.5	1865	10	Q81909	T7123.15 PROTEIN.	6.95e+00
39	91	5.4	257	1	Q9YAF9	257AA LONG HYPOTHETICA	9.28e+00
40	91	5.4	805	5	Q93719	F43G9.7 PROTEIN.	9.28e+00
41	91	5.4	833	4	Q75969	PROTEIN KINASE A BINDI	9.28e+00
42	91	5.4	1282	2	Q46348	COLLAGENASE PRECURSOR.	9.28e+00
43	91	5.4	1287	4	Q15468	SIL.	9.28e+00
44	90	5.3	388	2	Q53772	PUTATIVE OXIDOREDUCTAS	1.24e+01
45	90	5.3	463	10	Q42794	ASPARTATE AMINOTRANSFE	1.24e+01

ALIGNMENTS

RESULT	1	PRELIMINARY;	PRT;	44 AA.
ID	Q88417			
AC	Q88417			
DT	01-NOV-1998 (TReMBLrel. 08, Created)			
DT	01-NOV-1998 (TReMBLrel. 08, Last sequence update)			
DT	01-NOV-1998 (TReMBLrel. 08, Last annotation update)			
DE	HYPOTHETICAL 5.0 KD PROTEIN (FRAGMENT).			
OS	Mus musculus (Mouse)			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;			
OC	Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	CABIN D.E., MCKEE-JOHNSON J.W., MATEJIC L.E., WILTSHIRE T.,			
RA	MTAAVDET A., HUO Y.K., KORENBERG J.R., REEVES R.H.;			
RT	"Physical and Comparative Mapping of Distal Mouse Chromosome 16."			
RL	Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AF045953; AAC21456.1; -			
KW	Hypothetical protein.			
FT	NON_TER 1			
FT	NON_TER 44			
SQ	SEQUENCE 44 AA; 5010 MW; 296FALDE CRC32;			

Query Match 6.5%; Score 110; DB 11; Length 44;
Best Local Similarity 34.3%; Pred. No. 2.54e-02;
Matches 12; Conservative 11; Mismatches 12; Indels 0; Gaps 0;

Db	1	RLKAAKDAIEALGSKKEIKMKFRSSVFWAAGP 35
QY	152	KMDESRKLFSDLGSSYAKQLGFRDSWVFAGKDL 186
		::: :::: : : :::::

RESULT	2	PRELIMINARY;	PRT;	3190 AA.
ID	Q01368			
AC	Q01368			
DT	01-JUL-1997 (TReMBLrel. 04, Created)			
DT	01-JUL-1997 (TReMBLrel. 04, Last sequence update)			
DT	01-NOV-1999 (TReMBLrel. 12, Last annotation update)			
DE	CREB-BINDING PROTEIN HOMOLOG.			

OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.

```

RX MEDLINE: 97263578.
RA AKIMARU H., CHEN Y., DAI P., HOU D.X., NONAKA M., SMOLIK S.M.,
RA ARMSTRONG S., GOODMAN R.H., ISHII S.;
RT "Drosophila CBP is a co-activator of cubitus interruptus in hedgehog
RT signalling.";
RL Nature 386:735-738(1997).
DR EMBL: U88570; AAB53050.1; -.
DR HSSP: P03622; 2IFO.
DR FLYBASE: FBgn0015624; neJ.
DR PFAM: PF00439; bromodomain; 1.
DR PFAM: PF00569; Z2; 1.
DR PRINTS: PR00503; BROMODOMAIN.
SQ SEQUENCE 3190 AA; 331879 MW; A44A25BF CRC32;

Query Match 6.5%; Score 110; DB 5; Length 3190;
Best Local Similarity 33.8%; Pred. No. 2.54e-02;
Matches 24; Conservative 14; Mismatches 30; Indels 3; Gaps 3;

Db 181 GGMQVVGTQMGKVMSPMS-ITSNNGNMGNAIPGMTIAQNLGNMVLTVNSVGGMGGM 239
: ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | :
QY 71 SGAANVVGPTMCFEDRMINSVKNVGRGLN-IALVNGTIGAVLGOKAFDVMYSGDVH-L 128
: ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | :
Db 240 VNHKLQOPGGG 250
: ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | :
QY 129 VKFLKEIPGA 139

RESULT 3
ID Q94526 PRELIMINARY; PRT; 1001 AA.
AC Q94526;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-JAN-1999 (TREMBLrel. 09, Last sequence update)
DT 01-MAY-1999 (TREMBLrel. 10, Last annotation update)
DE TWO P DOMAIN POTASSIUM CHANNEL ORK1.
OS ORK1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN AND MUSCLES OF ADULTS;
RA GOLDSTEIN S.A., PRICE L.A., ROSENTHAL D.N., PAUSCH M.H.;
RX MEDLINE: 97075152.
RT "ORK1, a potassium-selective leak channel with two pore domains cloned
RT from Drosophila melanogaster by expression in Saccharomyces
RT cerevisiae.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:13256-13261(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN AND MUSCLES OF ADULTS;
RA GOLDSTEIN S.A.N., PRICE L.A., ROSENTHAL D.N., PAUSCH M.H.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN AND MUSCLES OF ADULTS;
RA GOLDSTEIN S.A.N., PRICE L.A., ROSENTHAL D.N., PAUSCH M.H.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: U55321; AAC69250.1; -.
DR FLYBASE: FBgn0017561; ORK1.
KW Ionic channel.
SQ SEQUENCE 1001 AA; 109290 MW; A57C4B04 CRC32;

Query Match 6.4%; Score 108; DB 5; Length 1001;
Best Local Similarity 33.3%; Pred. No. 4.90e-02;
Matches 20; Conservative 14; Mismatches 20; Indels 6; Gaps 6;

Db 97 FFAFTVCSTVGNTISPTT-FAGRMIMTAY-SVIGIPVNGILFAGL-GEYFG-RFFEAIY 152
: ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | :
QY 64 YFAFKICSCAA-NVVGPTMCFEDRMINSVKNVGRGLNIALVNGTIGAVLGOKAFD-MY 121
: ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | :

RESULT 4

```

```

ID O45691 PRELIMINARY; PRT; 358 AA.
AC O45691;
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)
DT 01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE MOES-2 PROTEIN.
GN MOES.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA DOBSON R.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SMALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
DR EMBL: Z93385; CAB07637.1; -.
DR PFAM: PF01018; GTP10BG.
DR PRINTS: PR00326; GTP10BG.
SQ SEQUENCE 358 AA; 39356 MW; E44EC56B CRC32;

Query Match 6.1%; Score 103; DB 5; Length 358;
Best Local Similarity 37.3%; Pred. No. 2.44e-01;
Matches 25; Conservative 11; Mismatches 27; Indels 4; Gaps 4;

Db 258 ISFLKHIERCESLWVLDYS-TGS-LTDQVKMLRVELEG-YQKGLGDRASTIVINKIDLS 314
: ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | :
QY 129 VKFLKEIPG-GALVLVASYDDPGTKMDESKRLFDLGSSTYAKOLGFRDSWVFIGAKDLR 187
: ||| | : ||| | : ||| | : ||| | : ||| | : ||| | : ||| | :
Db 315 GKSPREE 321
: ||| | :
QY 188 GKSPFEQ 194

RESULT 5
ID O29937 PRELIMINARY; PRT; 253 AA.
AC O29937;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-AUG-1998 (TREMBLrel. 07, Last annotation update)
DE HYPOTHETICAL 30.2 KD PROTEIN.
GN AF0307.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VG-16 / DSM 4304 / ATCC 49558;
RX MEDLINE: 98049343.
RA KLENK H.-P., CLAYTON R.A., TOMB J.-F., WHITE O., NELSON K.E.,
RA KETCHUM K.A., DODSON R.J., GWINN M., HICKEY E.K., PETERSON J.D.,
RA RICHARDSON D.L., KERLAVAGE A.R., GRAHAM D.E., KYRPIDES N.C.,
RA FLEISCHMANN R.D., QUACKENBUSH J., LEE N.H., SUTTON G.G., GILL S.,
RA KIRKNESS E.F., DOUGHERTY B.A., MCKENNEY K., ADAMS M.D., LOFTUS B.,
RA PETERSON S., REICH C.T., MCNEIL L.K., BADGER J.H., GLODEK A., ZHOU L.,
RA OVERBECK R., GOCAYNE J.D., WEIDMAN J.F., MCDONALD L., UTTERBACK T.,
RA COTTON M.D., SPRIGGS T., ARTIACH P., RAINE B.P., SYKES S.M.,
RA SADOW P.W., D'ANDREA K.P., BOWMAN C., FUJII C., GARLAND S.A.,
RA MASON T.M., OLSEN G.J., FRASER C.M., SMITH H.O., WOESE C.R.,

```

OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC	Ephydroidea; Drosophilidae; Drosophila.
[L]	
RN	SEQUENCE FROM N.A.
RC	TISSUE=EMBRYONIC AND IMAGINAL DISC;
RA	GREGORY S.L., BROWN N.H.;
RT	"kakapo," A Gene Required for Adhesion Between and Within Cell Layers
RT	In drosophila, Encodes a Large Cytoskeletal Linker Protein Related to
RL	Plectin and Dysmorphin."
RL	Submitted (Oct-1998) to the EMBL/GenBank/DDBJ databases.
DR	EMBL; AJ011924; CAAO9869.1; -.
DR	HSSP; QO1082; IAA2.
DR	PFAM; PF00307; SH3: 2.
DR	PFAM; PF00018; SH3: 1.
DR	PFAM; PF00435; spectrin; 15.
FT	NON_TER 2396 2396
SQ	SEQUENCE 2396 AA; 2D316313 CRC32;
	Query Match 6.0%; Score 102; DB 5; Length 2396;
	Best Local Similarity 26.5%; Pred.No. 3.35e-01;
Matches	18; Conservative 22; Mismatches 25; Indels 3; Gaps 3;
Dd	962 IPGACLLPPP-DQEIDAERKLRFDSRVALWQKHHLRQNMFATIRVVK-G-WDFD 1019 : : : : : : : : : : : :
Qy	135 IPPGALLVLVSDDPGTKMDESKRLF-S-DLGSSYAKQLGFSDSWVFIGAKDLRGKSPFE 193 : : : : : : : : : : : :
Dd	1020 QLTAWGPE 1027 :
Qy	194 QFLANSFD 201 :
RESULT	8 PRELIMINARY; PRT; 510 AA.
ID	O963B7 AC O963B7;
DT	01-MAY-1999 (TREMBLrel. 10, Created)
DT	01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT	01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE	INOSINE-5'-MONOPHOSPHATE DEHYDROGENASE.
OS	Plasmodium falciparum.
OC	Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium. [L]
RN	SEQUENCE FROM N.A.
RA	PROSISE G.L., JAMES A.A.; LUBCKE H.;
RL	Submitted (Nov-1997) to the EMBL/GenBank/DDBJ databases.
DR	EMBL; AF035679; AAD10256.1; -.
DR	HSSP; P50097; IAK5.
DR	PROSITE; PS00487; IMP_DH_GMP_RED; 1.
SQ	SEQUENCE 510 AA; 56148 MW; 196FA02C CRC32;
	Query Match 5.9%; Score 100; DB 5; Length 510;
	Best Local Similarity 21.2%; Pred.No. 6.25e-01;
Matches	18; Conservative 26; Mismatches 38; Indels 3; Gaps 3;
Dd	14 VMSVTVDVII-CMPGYIDFALSDDLTLNNMTDNITLKTPVISSPMDTVTHGHKSIALLS 72 : : : : : : : : : : : : : : : : :
Qy	49 VKKYKCGLIKPCPNYPFAFKICS-GAANVGPTWCFFEDRMISPVKNVGRGNIAL-VN 106 : : : : : : : : : : : : : : : : :
Dd	73 GELGVIHNNMSIEKOIIEVVKKVRf 97 : : : : : : : : : : : : : : : : :
Qy	107 GTTGAIVGGKAQAFMYSGDVMLHVKF 131 : : : : : : : : : : : : : : : : :
RESULT	9 PRELIMINARY; PRT; 531 AA.
ID	Q9XS55 AC Q9XS55;
DT	01-NOV-1999 (TREMBLrel. 12, Created)
DT	01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT	01-NOV-1999 (TREMBLrel. 12, Last annotation update)
DE	UDP--GLUCURONOSYLTRANSFERASE.
GN	SHUGTIIA6.
OS	Ovis aries (Sheep).
OC	Eukarya; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC	Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovioidea; Bovidae;

```
OC Caprinae; Ovis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SAFORK;
RX MEDLINE: 99208485.
RA KOBAYASHI T., TATANO A., YOKOTA H., ONAGA T., WATANABE T., YUASA A.;
RT "Small intestinal UDP-glucuronosyltransferase sheUGT1A07: partial
purification and cDNA cloning from sheep small intestine.";
RL Arch. Biochem. Biophys. 364:143-152(1999).
DR EMBL: AB018477; BAA77457.1; -.
DR PROSITE: PS00375; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 531 AA; 60823 MW; DC5D6014 CRC32;

Query Match 5.9%; Score 99; DB 6; Length 531;
Best Local Similarity 23.3%; Pred. No. 8.52e-01;
Matches 27; Conservative 36; Mismatches 43; Indels 10; Gaps 10;

Db 345 KNTKLKVLKLPNDLLGQPKT-R-AFITH-SGSHGVYEGICNGVPMVMP-L-FGDQMDNA 399
QY 32 KTIRLPRWLASPTKEIQVKYKCGLIKPCPANYFAFK-ICSGAANVVGPTMCFEDRIMS 90
Db 400 ERMETRAGITLVNLEMSGGLENALKAVINEKSYKENIMRLSLRHKDRPIEPLD 455
QY 91 PVKNNVGRGLNIALVNGTTGAVLQ-RA-F-DM-YSGDVHVLKFLKEIPGGALVL 142

RESULT 10
ID Q9XS56 PRELIMINARY; PRT; 532 AA.
AC Q9XS56;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DE UDP-GLUCURONOSYLTRANSFERASE.
GN SHEUGT1A07.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Caprinae; Ovis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SAFORK;
RX MEDLINE: 99208485.
RA KOBAYASHI T., TATANO A., YOKOTA H., ONAGA T., WATANABE T., YUASA A.;
RT "Small intestinal UDP-glucuronosyltransferase sheUGT1A07: partial
purification and cDNA cloning from sheep small intestine.";
RL Arch. Biochem. Biophys. 364:143-152(1999).
DR EMBL: AB018478; BAA77458.1; -.
DR PROSITE: PS00375; UDPGT; 1.
KW Transferase.
SQ SEQUENCE 532 AA; 60718 MW; F0E85701 CRC32;

Query Match 5.9%; Score 99; DB 6; Length 532;
Best Local Similarity 23.3%; Pred. No. 8.52e-01;
Matches 27; Conservative 36; Mismatches 43; Indels 10; Gaps 10;

Db 346 KNTKLKVLKLPNDLLGQPKT-R-AFITH-SGSHGVYEGICNGVPMVMP-L-FGDQMDNA 400
QY 32 KTIRLPRWLASPTKEIQVKYKCGLIKPCPANYFAFK-ICSGAANVVGPTMCFEDRIMS 90
Db 401 ERMETRAGITLVNLEMSGGLENALKAVINEKSYKENIMRLSLRHKDRPIEPLD 456
QY 91 PVKNNVGRGLNIALVNGTTGAVLQ-RA-F-DM-YSGDVHVLKFLKEIPGGALVL 142

RESULT 11
ID O82844 PRELIMINARY; PRT; 386 AA.
AC O82844;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DE XYLOSE REPRESSOR.
GN XYLR.
```

```
OS Tetragenococcus halophilus (Pediococcus halophilus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-I-13;
RX MEDLINE: 98320588.
RA TAKEDA Y., TAKASE K., YAMATO I., ABE K.;
RT "Sequencing and characterization of the xyl operon of a gram-positive
bacterium, Tetragenococcus halophila.";
RL Appl. Environ. Microbiol. 64:2513-2519(1998).
DR EMBL: AB009593; BAA31870.1; -.
DR PROSITE: PS01125; ROK; 1.
DR PFAM: PF00480; ROK; 1.
SQ SEQUENCE 386 AA; 43426 MW; B2B31CBB CRC32;

Query Match 5.8%; Score 98; DB 2; Length 386;
Best Local Similarity 29.6%; Pred. No. 1.16e+00;
Matches 21; Conservative 19; Mismatches 29; Indels 2; Gaps 2;

Db 290 EVIQLKNNKALISIGNNIVSMYDPTVINNELYREIPDLINVIEKHLTRNNRNVTI 349
QY 124 DVMLHVKF-LKEIPGALVLVASYDDPCTKWNDESRKLFSDLGSSYAKQLGFRDSW-VFI 181
Db 350 KNTSLEDKTTL 360
QY 182 GAKDLRGKSPF 192

RESULT 12
ID Q921K1 PRELIMINARY; PRT; 396 AA.
AC Q921K1;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE HSI BINDING PROTEIN 3.
GN HSI1BP3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C;
RA TAKEMOTO Y., FURUTA M., SATO M., HASHIMOTO Y.;
RT "Isolation and characterization of a novel HSI SH3 domain binding
protein, HSI1BP3.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ132192; CAAL0600.1; -.
SQ SEQUENCE 396 AA; 43769 MW; ODCA7DDE CRC32;

Query Match 5.8%; Score 97; DB 11; Length 396;
Best Local Similarity 32.6%; Pred. No. 1.57e+00;
Matches 14; Conservative 11; Mismatches 17; Indels 1; Gaps 1;

Db 98 LFVGESDIRERAMFDELLRCVSKDAQLAGSPELLEFLGTRAP 140
QY 179 VFIGAKDLRGKSP-FEQLKNSPTNKNYEGWPELLEMEGCMPP 220

RESULT 13
ID O52072 PRELIMINARY; PRT; 561 AA.
AC O52072;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE PHA SYNTHASE.
GN PHAC.
OS Nocardia corallina.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Nocardiaceae; Nocardia.
RN [1]
RP SEQUENCE FROM N.A.
RA HALL B., BALDWIN J., RHIE H.G., DENNIS D.;
```


Search completed: Wed May 31 11:05:21 2000
Job time : 26 secs.

mpsrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 31 11:04:26 2000; MasPar time 8.90 Seconds
762.806 Million cell updates/sec
Tabular output not generated.

```
>US-09-186-342-2
Title:
Description: (1-223) from US09186342.pap
Perfect Score: 1686
Sequence: 1 MRVSGVRLRLALIFAIVTW.....KYEGWPELLEMEGCMNPKPF 223
```

Scoring table: PAM 150
Gap 11

Searched: 83857 seqs, 30454973 residues

Post-processing: Minimum Match 08 Listing first 45 summaries

Database: swiss-prot38
1:swissprot

```

statistics:
  Mean 47.338:  Variance 81.778:  scale 0.579

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			DB	ID	Description	Pred. No.
	Score	Match	Length				
1	908	53	9	230	1	219_HUMAN	2.11e-179
2	855	50	7	125	1	G786_HUMAN	9.26e-167
3	658	35	0	115	1	OTTH_MOUSE	3.13e-120
4	104	6	2	312	1	DHBK_MOUSE	6.70e-07
5	96	5	7	522	1	TSAM_RICTS	8.83e-01
6	96	5	7	532	1	TSAK_RICTS	8.83e-01
7	95	5	6	124	1	YRFE_YEAST	1.21e+00
8	95	5	6	224	1	TRPE_YEAST	1.21e+00
9	95	5	6	349	1	FML2_MACMU	1.21e+00
10	95	5	6	885	1	CHS3_EXODE	1.21e+00
11	92	5	5	215	1	SOMA_TRIVU	3.01e+00
12	91	5	4	166	1	TH12_BOVIN	4.07e+00
13	91	5	4	287	1	DMSC_ECOLI	4.07e+00
14	91	5	4	382	1	YAC3_SCHPO	4.07e+00
15	91	5	4	563	1	ACL_ORYSA	4.07e+00
16	89	5	3	104	1	ARTA_ECOLI	7.34e+00
17	90	5	3	137	1	VYFS_ECOLI	5.47e+00
18	89	5	3	159	1	VFPL_FOWP1	7.34e+00
19	89	5	3	204	1	SOMA_SERQU	7.34e+00
20	90	5	3	256	1	MOVP_TWGWU	5.47e+00
21	89	5	3	361	1	SHAK_DRDME	7.34e+00
22	90	5	3	410	1	CREA_BACSP	5.47e+00
23	90	5	3	428	1	YM14_CAEEL	5.47e+00

[illegible]

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
CC DR EMBL: AF064635; AAC16885.1; -.
CC DR PROSITE: PS00061; ADH_SHORT; 1.
CC DR PFWA; PF00106; adh_short; 1.
CC KW Steroid biosynthesis; Oxidoreductase; NADP; Multigene family.
CC FT NP_BIND 50 79 NADP (BY SIMILARITY).
CC FT ACT_SITE 202 202 BY SIMILARITY.
CC SQ SEQUENCE 312 AA; 34742 MW; 7D255188FALF8DDB CRC64;

Query Match 6.2%; Score 104; DB 1; Length 312;
Best Local Similarity 34.0%; Pred. No. 6.70e-02;
Matches 16; Conservative 14; Mismatches 13; Indels 4; Gaps 4;

Db 268 RTGVYHISGINSIMPRVMVFKIIMGFS-KSLR-NRYLKKRKN 312

QY 2 RVSG-VLR-LLALIFAIVTWEIFRYSFMSFKTIRLPWLASPTKE 46

RESULT 5

ID TSAW_RICTS STANDARD; PRT; 522 AA.
AC P37919;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DE 56 KD TYPE-SPECIFIC ANTIGEN PRECURSOR (TSA) (56 KD SCRUB TYPHUS
DE ANTIGEN) (TSA56) (TSK56).
OS Rickettsia tsutsugamushi.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Orientia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KAWASAKI;
RX MEDLINE: 92316959.
RA Ohashi N., Nashimoto H., Ikeda H., Tamura A.;
RT "Diversity of immunodominant 56-kDa type-specific antigen (TSA) of
RT Rickettsia tsutsugamushi. Sequence and comparative analyses of the
RT genes encoding TSA homologues from four antigenic variants.*";
RL J. Biol. Chem. 267:12735(1992).
CC -1- FUNCTION: MAY BE AN ADHERENT FACTOR FOR RICKETTSIAL ADSORPTION TO
CC THE HOST-CELL SURFACE AND A DETERMINANT OF VIRULENCE OF INDIVIDUAL
CC RICKETTSIAL STRAIN. IT IS THE MAJOR OUTER MEMBRANE PROTEIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: M63383; AAA26398.1; -.
CC PIR: B42804; B42804.
CC KW Antigen; Signal; Transmembrane; Virulence.
CC FT SIGNAL 1 22 POTENTIAL.
CC FT CHAIN 23 522 56 KD TYPE-SPECIFIC ANTIGEN.
CC FT DOMAIN 324 353 ALA/GLN-RICH.
CC FT TRANSMEM 67 87 POTENTIAL.
CC FT TRANSMEM 470 490 POTENTIAL.
CC SQ SEQUENCE 522 AA; 55971 MW; 64C80E3B1B5A2EB6 CRC64;

Query Match 5.7%; Score 96; DB 1; Length 522;
Best Local Similarity 31.0%; Pred. No. 8.83e-01;
Matches 18; Conservative 20; Mismatches 17; Indels 3; Gaps 3;

Db 252 VLSDKITQIYS-DIRQFAKIANIEVPGAPLPNSASVEIQTKMQELN-DVLEELRESF 307

QY 112 VLGGKAFDMYSGDVHVLVKFLK-BIPGGALVLVASYDDPGTKMNDESKRFLSGLSSY 168

RESULT 6

ID TSAK_RICTS STANDARD; PRT; 532 AA.
AC P37915;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last annotation update)
DE 56 KD TYPE-SPECIFIC ANTIGEN PRECURSOR (TSA) (56 KD SCRUB TYPHUS
DE ANTIGEN) (TSA56) (TSK56).
OS Rickettsia tsutsugamushi.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsieae; Orientia.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-KARP;
RX MEDLINE: 90307205.
RA Stover C.K., Marana D.P., Carter J.M., Roe B.A., Mardis E.,
RA Oaks E.V.;
RT "The 56-kilodalton major protein antigen of Rickettsia tsutsugamushi:
RT molecular cloning and sequence analysis of the sta56 gene and precise
RT identification of a strain-specific epitope.*";
RL Infect. Immun. 58:2076-2084(1990).
CC -1- FUNCTION: MAY BE AN ADHERENT FACTOR FOR RICKETTSIAL ADSORPTION TO
CC THE HOST-CELL SURFACE AND A DETERMINANT OF VIRULENCE OF INDIVIDUAL
CC RICKETTSIAL STRAIN. IT IS THE MAJOR OUTER MEMBRANE PROTEIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: M33004; AAA26391.1; -.
CC DR Antigen; Signal; Transmembrane; Virulence.
CC KW SIGNAL 1 22 POTENTIAL.
CC FT CHAIN 23 532 56 KD TYPE-SPECIFIC ANTIGEN.
CC FT DOMAIN 339 364 ALA/GLN-RICH.
CC FT TRANSMEM 67 87 POTENTIAL.
CC FT TRANSMEM 480 500 POTENTIAL.
CC SQ SEQUENCE 532 AA; 56785 MW; B63F4BE5CF917638 CRC64;

Query Match 5.7%; Score 96; DB 1; Length 532;
Best Local Similarity 30.2%; Pred. No. 8.83e-01;
Matches 19; Conservative 17; Mismatches 24; Indels 3; Gaps 3;

Db 263 VLSDKITQIYS-DIKHLADIAGIDVPDTSLPNSASVEIQTKMQELN-DLLEELRESFDG 320

QY 112 VLGGKAFDMYSGDVHVLVKFL-KEIPGGALVLVASYDDPGTKMNDESKRFLSGLSSYAK 170

Db 321 YLG 323

QY 171 QLG 173

RESULT 7

ID Y48A_HAEIN STANDARD; PRT; 124 AA.
AC O86225;
DT 15-FEB-2000 (Rel. 39, Created)
DT 15-FEB-2000 (Rel. 39, Last sequence update)
DT 15-FEB-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL PROTEIN HI0485.1.
GN HI0485.1.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20;
RX MEDLINE: 95350630.
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,

RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.W., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus
 RT Influenzae Rd.";
 RL Science 269:496-512(1995).
 RN [2]
 RP IDENTIFICATION
 RA White O., Clayton R.A., Kerlavage A.R., Fleischmann R.D., Peterson J.,
 RA Hickey E., Dodson R., Gwinn M.;
 RA Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U32730; AAC22145.1; -;
 DR TIGR; H10485.1; -;
 DR KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 14 34 POTENTIAL.
 FT TRANSMEM 41 61 POTENTIAL.
 FT TRANSMEM 85 105 POTENTIAL.
 SQ SEQUENCE 124 AA; 14546 MW; E4D867EDA4FB4C3E CRC64;
 Query Match 5.6%; Score 95; DB 1; Length 124;
 Best Local Similarity 24.1%; Pred. No. 1.21e+00;
 Matches 7; Conservative 14; Mismatches 7; Indels 1; Gaps 1;
 Db 75 RBAKFIITWVFIIRKWLVINFIAT 103
 QY 2 RVSGVLRLALIFAIVTT-WMFIRSYMSE 29
 RESULT 8
 ID TRPE_YEAST STANDARD; PRT; 224 AA.
 AC P00912;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 15-FEB-2000 (Rel. 39, Last annotation update)
 DE N-(5'-PHOSPHORIBOSYL)-ANTHRANILATE ISOMERASE (EC 5.3.1.24) (PRAI).
 GN TRP1 OR YDR007W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
 OC Saccharomycetaceae; Saccharomycetes.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 80225748.
 RA Tschumper G., Carbon J.;
 RT "Sequence of a yeast DNA fragment containing a chromosomal replicator
 RT and the TRP1 gene.";
 RL Gene 10:157-166(1980).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Brunelli J.P., Hilgers L., Pall M.L.;
 RL Submitted (XXX-1991) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 68-224 FROM N.A.
 RC STRAIN=S288C / AB972;
 RA Murphy L., Richards C., Gentles S., Harris D., Barrell B.G.,
 RA Rajandream M.A.;
 RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE OF 206-224 FROM N.A.
 RX MEDLINE; 87065038.

RA Snyder M., Buchman A.R., Davis R.W.;
 RT "Bent DNA at a yeast autonomously replicating sequence.";
 RL Nature 324:87-89(1986).
 CC -|- CATALYTIC ACTIVITY: N-(5'-PHOSPHORIBOSYL)-ANTHRANILATE =
 CC 1-(2-CARBOXYPHENYLAMINO)-1-DESOXY-RIBULOSE 5-PHOSPHATE.
 CC -|- PATHWAY: THIRD STEP IN TRYPTOPHAN BIOSYNTHESIS.
 CC -|- SIMILARITY: BELONGS TO THE TRP FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; V01341; CAA24634.1; -;
 DR EMBL; J01374; AAA88758.1; -;
 DR EMBL; M74015; AAA72097.1; -;
 DR EMBL; Z48008; CAA8068.1; -;
 DR EMBL; M30386; AAA18406.1; -;
 DR PIR; A01135; ISBYN.
 DR HSSP; Q56320; INSJ.
 DR SGD; L0002352; TRP1.
 DR PFAM; PF00697; PRAI; 1.
 DR KW Isomerase; Tryptophan biosynthesis.
 FT CONFLICT 58 58 T->S (IN REF. 2).
 SQ SEQUENCE 224 AA; 24158 MW; CEEDB0EEF4F97911 CRC64;
 Query Match 5.6%; Score 95; DB 1; Length 224;
 Best Local Similarity 25.8%; Pred. No. 1.21e+00;
 Matches 25; Conservative 21; Mismatches 47; Indels 4; Gaps 4;
 Db 61 KAYKNSGTPKYLGVTRNPQKEDVLALVNDYGDIVQLHGDSWQBYQFGLGLPVTKRL 120
 QY 116 KAFDMSYGVNHLVKFLKEIP-GGALVLVASYDDPGTKMN-DESRKLFSD-LGSSVAKQL 172
 Db 121 VFPRDCNLLSAASQKPHSFIFPLFDSAGGTGELLDW 157
 QY 173 GF-RDSNVFVGAKDLRGKSPFEQFLKNSPDTNKYEGW 208
 RESULT 9
 ID FML2_MACMU STANDARD; PRT; 349 AA.
 AC P79191;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE N-FORMYL PEPTIDE RECEPTOR-LIKE 2 RECEPTOR (FRAGMENT).
 GN FPRL2.
 OS Macaca mulatta (Rhesus macaque).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 OC Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecinae;
 OC Macaca.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 96421539.
 RA Alvarez V., Coto E., Sehen F., Gouzalek-Koces S., Lopez-Larrea C.;
 RT "Molecular evolution of the N-formyl peptide and csa receptors in
 RT non-human primates.";
 RL Immunogenetics 44:446-452(1996).
 CC -|- FUNCTION: LOW AFFINITY RECEPTOR TO N-FORMYL-METHIONYL PEPTIDES,
 CC WHICH ARE POWERFUL NEUTROPHILS CHEMOTACTIC FACTORS. BINDING OF
 CC FMLP TO THE RECEPTOR CAUSES ACTIVATION OF NEUTROPHILS. THIS
 CC RESPONSE IS MEDIATED VIA A G-PROTEIN THAT ACTIVATES A
 CC PHOSPHATIDYLINOSITOL-CALCIUM SECOND MESSENGER SYSTEM.
 CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
 CC -|- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

DR EMBL; X97740; CAA66324.1; -.
 DR GCRDB; GCR.1739; -.
 DR PROSITE; PS00237; G-PROTEIN_RECEPTOR; 1.
 DR PFAM; PF00001; 7tm_1; 2.
 KW G-protein coupled receptor; Transmembrane; Glycoprotein;
 KW Chemotaxis.
 FT DOMAIN 1 27 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 28 50 1 (POTENTIAL).
 FT DOMAIN 51 61 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 62 83 2 (POTENTIAL).
 FT DOMAIN 84 100 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 101 121 3 (POTENTIAL).
 FT DOMAIN 122 140 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 141 162 4 (POTENTIAL).
 FT DOMAIN 163 205 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 206 226 5 (POTENTIAL).
 FT DOMAIN 227 242 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 243 266 6 (POTENTIAL).
 FT DOMAIN 267 286 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 287 306 7 (POTENTIAL).
 FT DOMAIN 307 >349 CYTOPLASMIC (POTENTIAL).
 FT CARBOHYD 4 4 POTENTIAL.
 FT CARBOHYD 10 10 POTENTIAL.
 FT DISULFID 98 176 POTENTIAL.
 FT NON_TER 349 349
 SQ SEQUENCE 349 AA; 39531 MW; C0C92A9E5CDE250C CRC64;

Query Match 5.6%; Score 95; DB 1; Length 349;
 Best Local Similarity 28.4%; Pred. NO. 1.21e+00;
 Matches 29; Conservative 27; Mismatches 37; Indels 9; Gaps 6;
 Db 231 KRMTKSRPLH-IFTAVASFFICWF-PYELTGILMAVWL-----KEILNGYKILVLI 284
 QY 1 MRYSGVLRLLALFAITVMTFIRSYMSFMTIRLPWLASPTKEIQVK-KYKGLIKP 59
 Db 285 NPTSSLAFFNSCLNPLYVFMGHNFQERLIRS-LPTSLEAL 325
 QY 60 CPANYEAF-KICSGAANVVGPTMCFEDRMIMSPVKNVGRGL 100

RESULT 10
 ID CHS3_EXODE STANDARD; PRT; 885 AA.
 AC P30602; O74678;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE CHITIN SYNTHASE 3 (EC 2.4.1.16) (CHITIN-UDP ACETYL-GLUCOSAMINYL
 DE TRANSFERASE 3) (CLASS-III CHITIN SYNTHASE 3).
 GN CHS3.
 OS Exophiala dermatitidis (Wangiella dermatitidis).
 OC Eukaryota; Fungi; Ascomycota; Chaetothiales; Herpotrichiellaceae;
 OC anamorphic Herpotrichiellaceae; Exophiala.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=8656;
 RA Wang Z., Graybill J.R., Szaniszlo P.J.;
 RT "Characterization and expression studies of WdCHS3, a gene that
 RT encodes a class III chitin synthase and contributes to virulence in
 RT Wangiella (Exophiala) dermatitidis";
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE OF 214-410 FROM N.A.
 RX MEDLINE; 92115692.
 RA Bowen A.R., Chen-Wu J.L., Momany M., Young R., Szaniszlo P.J.,
 RA Robbins P.W.;
 RT "Classification of fungal chitin synthases";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:519-523(1992).
 CC -1- FUNCTION: PLAYS A MAJOR ROLE IN CELL WALL BIOGENESIS.
 CC -1- CATALYTIC ACTIVITY: UDP-N-ACETYL-D-GLUCOSAMINE + [1,4-(N-ACETYL-

CC BETA-D-GLUCOSAMINYL)](N) = UDP + [1,4-(N-ACETYL-BETA-D-
 CC GLUCOSAMINYL)](N+1).
 CC -1- SUBCELLULAR LOCATION: PLASMA MEMBRANE-BOUND.
 CC -1- SIMILARITY: BELONGS TO THE CHITIN SYNTHASE FAMILY. SUBFAMILY CLASS
 CC III.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF053314; AAC35278.1; -.
 DR EMBL; M81907; AAA30336.1; -.
 DR PFAM; PF01644; Chitin_synth; 1.
 KW Transferase; Glycosyltransferase; Transmembrane; Cell wall;
 KW Multigene family.
 FT CONFLICT 215 215 Y -> M (IN REF. 2).
 FT CONFLICT 340 340 R -> S (IN REF. 2).
 SQ SEQUENCE 885 AA; 99422 MW; 48D00BEE408398F0 CRC64;
 Query Match 5.6%; Score 95; DB 1; Length 885;
 Best Local Similarity 28.6%; Pred. NO. 1.21e+00;
 Matches 14; Conservative 17; Mismatches 13; Indels 5; Gaps 4;
 Db 837 LVATWFSNALLAVATSDLSDRGFTSEPLRGPAISRPF-CGLLRPC 884
 QY 16 IVTWMFIRSYMSFMTIRLPW-LAS-PTK--EIQVKKYKGLRKC 60
 RESULT 11
 ID SOMA_TRIVU STANDARD; PRT; 215 AA.
 AC O62754;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE SOMATOTROPIN PRECURSOR (GROWTH HORMONE).
 GN GHI.
 OS Trichosurus vulpecula (Brush-tailed possum).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
 OC Metatheria; Diprotodontia; Phalangeridae; Trichosurus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE; 98325478.
 RA Saunders M.C., Deakin J., Harrison G.A., Curlewis J.D.;
 RT "cDNA cloning of growth hormone from the brushtail possum
 RT (Trichosurus vulpecula).";
 RL Gen. Comp. Endocrinol. 111:68-75(1998).
 CC -1- FUNCTION: GROWTH HORMONE PLAYS AN IMPORTANT ROLE IN GROWTH
 CC CONTROL.
 CC -1- SUBCELLULAR LOCATION: SECRETED.
 CC -1- SIMILARITY: BELONGS TO THE SOMATOTROPIN/PROLACTIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF052192; AAC08986.1; -.
 DR PRINTS; PR00836; SOMATOTROPIN.
 DR PROSITE; PS00266; SOMATOTROPIN_1; 1.
 DR PROSITE; PS00338; SOMATOTROPIN_2; 1.
 DR PFAM; PF00103; hormone; 1.
 KW Hormone; Pituitary; Signal.
 FT SIGNAL 1 25
 FT CHAIN 26 215
 FT DISULFID 77 188
 FT DISULFID 205 213
 BY SIMILARITY.
 BY SIMILARITY.
 BY SIMILARITY.

```

SQ SEQUENCE 215 AA; 24353 MW; F241085B4A7352D1 CRC64;

Query Match 5.5%; Score 92; DB 1; Length 215;
Best Local Similarity 32.2%; Pred. No. 3.01e+00;
Matches 28; Conservative 24; Mismatches 28; Indels 7; Gaps 7;

Db 111 LSPVQF-LSRVFNSLVFTSDRYV-EKLRDLLEG-IQALMGELEDGSGRGLVLTYYD 167
      :||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 89 MSPVKNNVGRGLNVLNGTTGAVLGOKAFDMYSGDVHVLVFLKRIPG-GALVILVASVD 147
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 168 KFTNLSRDEALKKNYG-LLSCFKDL 193
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 148 DPCTKM-NDESR-KLFSDLGSSYAKOL 172
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

RESULT 12
ID TH12_BOVIN STANDARD; PRT; 166 AA.
AC Q95108;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DE 15-FEB-2000 (Rel. 39, Last annotation update)
DE MITOCHONDRIAL THIREDOXIN PRECURSOR (MT-TRX).
GN TXN2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovinae; Bos.
[1]
RN RP SEQUENCE FROM N.A., AND SEQUENCE OF 60-106 AND 111-166.
RC TISSUE-ADRENAL CORTEX;
RX MEDLINE; 98028383.
RA Watabe S., Hiroi T., Yamamoto Y., Fujioka Y., Hasegawa H., Yago N.,
RA Takahashi S.Y.;
RT "SP-22 is a thioredoxin-dependent peroxide reductase in
RT mitochondria.";
RL Eur. J. Biochem. 249:52-60(1997).
CC -1- FUNCTION: POSSES A DITHIOL-REDUCING ACTIVITY. COOPERATES WITH SP-
CC 22 PROTEIN TO PROTECT GLUTAMINE SYNTHETASE AND OTHER PROTEINS FROM
CC FERROUS ION/DITHIOREITOL-MEDIATED DAMAGE. WITHOUT SP-22, THE
CC PROTEIN HAD NO PROTECTING ACTIVITY.
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC -1- SIMILARITY: BELONGS TO THE THIREDOXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D87741; BAA13447.1; -
CC PRINTS; PR00421; THIREDOXIN.
CC DR PROSITE; PS00194; THIREDOXIN; 1.
CC DR PFAM; PF00085; thiored; 1.
CC KW Redox-active center; Electron transport; Mitochondrion;
CC TRANSIT peptide.
CC FT TRANSIT 1 59 MITOCHONDRION.
CC FT CHAIN 60 166 MITOCHONDRIAL THIREDOXIN.
CC FT DISULFID 90 93 REDOX-ACTIVE (BY SIMILARITY).
CC SQ SEQUENCE 166 AA; 18416 MW; 5596F7DDA7E34D5F CRC64;

Query Match 5.4%; Score 91; DB 1; Length 166;
Best Local Similarity 19.2%; Pred. No. 4.07e+00;
Matches 15; Conservative 27; Mismatches 31; Indels 5; Gaps 5;

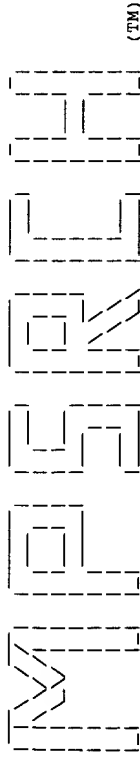
Db 1 MAQRLLRLRFTLSITSGKPSQSRNPASVALKTPQYSPGLTVTPSQARSIVTRVCST 60
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 17 VTTWVFIRSYMSFKTIR-LPRWLASPTKEIQVKYKCG-L-IKPCPA-NYFAFKICSG 72
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 61 TFNTOGDPD-FQDRVYNS 77
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
QY 73 AANVVGPTMCFEDRMIMS 90
      :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||

```

```

RESULT 13
ID DMSO_ECOLI STANDARD; PRT; 287 AA.
AC P18777;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE ANAEROBIC DIMETHYL SULFOXIDE REDUCTASE CHAIN C (DMSO REDUCTASE ANCHOR
DE SUBUNIT).
DE DMSO.
GN Escherichia coli.
OS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
[1]
RN RP SEQUENCE FROM N.A.
RC STRAIN-K12 / C600;
RX MEDLINE; 89096500.
RA Bilous P.T., Cole S.T., Anderson W.F., Weiner J.H.;
RT "Nucleotide sequence of the dmsABC operon encoding the anaerobic
RT dimethylsulphoxide reductase of Escherichia coli.";
RL Mol. Microbiol. 2:785-795(1988).
[2]
RN RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 97426617.
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner F.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
[3]
RN RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE; 97061202.
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Kimura K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampaio G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-Kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
[4]
RN RP TOPOLOGY.
RX MEDLINE; 93155163.
RA Weiner J.H., Shaw G., Turner R.J., Trieber C.A.;
RT "The topology of the anchor subunit of dimethyl sulfoxide reductase
RT of Escherichia coli.";
RL J. Biol. Chem. 268:3238-3244(1993).
CC -1- FUNCTION: TERMINAL REDUCTASE DURING ANAEROBIC GROWTH ON
CC VARIOUS SULFOXIDE AND N-OXIDE COMPOUNDS. THE C SUBUNIT ANCHORS
CC THE OTHER TWO SUBUNITS TO THE MEMBRANE AND STABILIZE THE CATALYTIC
CC SUBUNITS.
CC -1- SUBUNIT: THE COMPLEX CONSIST OF THREE SUBUNITS: DMSA, THE
CC REDUCTASE; DMSB, AN ELECTRON TRANSFER PROTEIN, AND DMSC, A
CC MEMBRANE ANCHOR PROTEIN.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J03412; AAA83845.1; -
CC EMBL; AE000191; AAC73982.1; -
CC EMBL; D90727; BAA35628.1; -
CC EMBL; D90728; BAA35631.1; -

```



Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm
Run on: Wed May 31 11:03:51 2000; MasPar time 14.02 Seconds
Tabular output not generated. 750.250 Million cell updates/sec

Title: >US-09-186-342-2
Description: (1-223) from US09186342.pep
Perfect Score: 1686
Sequence: 1 MRVSGVLRLLALIFAIVTW.....KYEGWPELLEMEGCMPPKPF 223

Scoring table: PAM 150
Gap 11

Searched: 142080 seqs, 47172406 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: pir62
1:pir1 2:pir2 3:pir3 4:pir4

Statistics: Mean 46.426; Variance 93.300; scale 0.498

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
1	908	53.9	230	2	I37095 gene 2.19 protein - h	2.76e+154
2	110	6.5	3190	2	T13828 CREB-binding protein	6.73e-02
3	108	6.4	1001	2	T13807 potassium channel pro	1.23e-01
4	101	6.0	253	2	C69288 hypothetical protein	9.50e-01
5	102	6.0	2285	2	T13715 kakapo gene protein i	7.13e-01
6	102	6.0	2396	2	T13714 kakapo gene protein -	7.13e-01
7	99	5.9	121	2	S69859 hypothetical protein	1.68e+00
8	97	5.8	383	2	G51533 probable n2,n2-dimeth	2.93e+00
9	97	5.8	1134	2	G69269 molybdopterin-binding	2.93e+00
10	96	5.7	269	2	T15631 hypothetical protein	3.87e+00
11	96	5.7	320	2	T15849 hypothetical protein	3.87e+00
12	96	5.7	389	2	T00697 glycylopeptide N-tetra	3.87e+00
13	96	5.7	522	2	B42804 immunodominant type-s	3.87e+00
14	95	5.6	224	1	ISBYN phosphoribosylanthran	5.09e+00
15	94	5.6	855	2	T10665 hypothetical protein	6.67e+00
16	92	5.5	291	2	T03886 hypothetical protein	1.14e+01
17	93	5.5	469	2	C70357 hypothetical protein	8.74e+00
18	92	5.5	509	2	S51348 hypothetical protein	1.14e+01
19	93	5.5	573	2	F70047 two-component sensor	8.74e+00
20	91	5.4	257	2	F72500 hypothetical protein	1.49e+01
21	91	5.4	287	2	S03787 probable dimethylsulf	1.49e+01
22	91	5.4	491	2	C75078 hypothetical protein	1.49e+01
23	91	5.4	563	1	JU0311 4-coumarate-CoA liga	1.49e+01

24	91	5.4	1282	2	JC4393 microbial collagenase	1.49e+01
25	91	5.4	1287	2	A41685 SIL protein - human	1.49e+01
26	89	5.3	104	1	BVECAA hypothetical protein	2.51e+01
27	90	5.3	137	2	H64860 hypothetical protein	1.94e+01
28	90	5.3	145	2	S07957 somatotropin precursor	2.51e+01
29	89	5.3	204	1	STFI 28.5K transport prote	1.94e+01
30	90	5.3	256	1	WMTMU2 copD homolog - Xantho	2.51e+01
31	89	5.3	307	2	D36868 NADH dehydrogenase su	2.51e+01
32	89	5.3	344	2	T14218 Pas (Passover) proteol	2.51e+01
33	89	5.3	361	2	A40734 probable oxidoreducta	1.94e+01
34	90	5.3	388	2	F70933 creatinase (EC 3.5.3.	1.94e+01
35	90	5.3	411	2	I39809 NAD(P)+ transhydrogen	1.94e+01
36	90	5.3	462	1	DEECXB aspartate transaminas	1.94e+01
37	90	5.3	463	2	S33528 6-phosphofructo-2-kin	1.94e+01
38	90	5.3	469	2	A40800 CDPdiacylglycerol--se	2.51e+01
39	89	5.3	521	2	T11166 peptidase transport pro	2.51e+01
40	89	5.3	579	2	T04378 peptidase homolog - L	2.51e+01
41	89	5.3	592	2	C70108 glycerol-3-phosphate	2.51e+01
42	89	5.3	649	2	S48379 probable cleavage and	1.94e+01
43	90	5.3	676	2	C72749 gibbon ape leukemia r	2.51e+01
44	89	5.3	679	2	I48084 leukemia virus recept	2.51e+01
45	89	5.3	679	2	I52822	

ALIGNMENTS

RESULT 1
ENTRY I37095 #type complete
TITLE gene 2.19 protein - human
ORGANISM #formal_name Homo sapiens
DATE 01-Nov-1996 #sequence_revision 01-Nov-1996 #text_change 28-Feb-1997
ACCESSIONS I37095
REFERENCE #authors
#journal Bione, S.; Tamanini, F.; Maestrini, E.; Tribioli, C.; Poustka, A.; Torri, G.; Rivella, S.; Toniolo, D.
#title Proc. Natl. Acad. Sci. U.S.A. (1993) 90:10977-10981
#cross-references EMBL:X87193; NID:g854081; PID:g854082
#cross-references MUID:94068527
#accession I37095
#status preliminary; translated from GB/EMBL/DBDJ
#molecule_type mRNA
#residues 1-230 #label RES
#cross-references EMBL:X87193; NID:g854081; PID:g854082
GENETICS
#gene 2.19
SUMMARY #length 230 #molecular-weight 25069 #checksum 6211

Query Match	53.9%	Score 908;	DB 2;	Length 230;
Best Local Similarity	53.3%;	Pred. No. 2.76e+154;		
Matches	121;	Conservative 56;	Mismatches 45;	Indels 5; Gaps 2;
Db	1	MRLAGPLRIVLVVGVTVVIVVILGGPGSGFPRIQQLFTSPSSVTAAPRKYKCG	60	
Qy	1	MRVSGVLRLLALIFAIVTWFTSRYSMSFKTI-RLPRWLASPTKEI-----QVKKYKCG	55	
Db	61	LPOCPPEHLAFRVVSGAANVIGPKICLEDKMLMSSVKDNGVGLNIALNVGSELIPA	120	
Qy	56	LIRKPCPANYFAFKISGAANVGVPTMCFEDRMIMSPVKNVGRGLNIALNVGTGAVLQ	115	
Db	121	RAFDWAGDVNDLLKTRPLHEGTLVFVASYDDPATKMNTEETKRLSELGSRNAKELAFR	180	
Qy	116	KAFDMYSGDVMHLVFKLKEIPGGALVLVASYDDPGTKMDESKRFLSDLGSSYAKQLGPR	175	
Db	181	DSWVFYGAQVQNKSPFEQHVKNKSKNSKYEGCPEALEMEGCIPIRRS	227	
Qy	176	DSWVFYGAQVQNKSPFEQHVKNKSKNSKYEGCPEALEMEGCIPIRRS	222	

RESULT 2
ENTRY T13828 #type complete
TITLE CREB-binding protein homolog - fruit fly (Drosophila)

ORGANISM	melanogaster)
DATE	#formal_name Drosophila melanogaster 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
ACCESSIONS	T13828
REFERENCE	Z17785
#authors	Akamaru, H.; Chen, Y.; Dai, P.; Hou, D.X.; Nonaka, M.; Smolik, S.M.; Armstrong, S.; Goodman, R.H.; Ishii, S.
#journal	Nature (1997) 386:735-738
#title	Drosophila Cbp is a co-activator of cubitus interruptus in hedgehog signalling.
#accession	T13828
#status	preliminary; translated from GB/EMBL/DBJ
#molecule_type	mRNA
#residues	1-3190 ##label AKI
#cross-references	EMBL:U88570; NID:g1916929; PID:g1916930; PIDN:AAB53050.1
GENETICS	
#map_position x	
SUMMARY	#length 3190 #molecular-weight 331879 #checksum 2759
Query Match	6.5%; Score 110; DB 2; Length 3190;
Best Local Similarity	33.8%; Pred. No. 6.73e-02;
Matches	24; Conservative 14; Mismatches 30; Indels 3; Gaps 3;
Db	181 GGMQVVYGTGMVNSPMPS-1SNNNGNMNAIPGMNTIAQGNLGNKVLTVNSVGGMGGM 239
Qy	71 SGAANVVGPTMCFEDRMISPVKNVGRGLN-IALVNGTTGAVLGOKAPDMYSGDVVH-L 128
Db	240 VNHLKQPPGG 250
Qy	129 VKFLKEIPGGA 139
RESULT	3
ENTRY	T13807
TITLE	#type complete potassium channel protein - fruit fly (Drosophila melanogaster)
ORGANISM	#formal_name Drosophila melanogaster
DATE	13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 13-Aug-1999
ACCESSIONS	T13807
REFERENCE	Z17770
#authors	Goldstein, S.A.; Price, L.A.; Rosenthal, D.N.; Pausch, M.H.
#journal	Proc. Natl. Acad. Sci. U.S.A. (1996) 93:13256-13261
#title	ORK1, a potassium-selective leak channel with two pore domains cloned from Drosophila melanogaster by expression in Saccharomyces cerevisiae.
#cross-references	MUID:97075152
#accession	T13807
#status	preliminary; translated from GB/EMBL/DBJ
#molecule_type	mRNA
#residues	1-1001 ##label GOL
#cross-references	EMBL:U55321; NID:g3808067; PID:g3808068; PIDN:AAC69250.1
GENETICS	
#gene	ORK1
#map_position 1	
SUMMARY	#length 1001 #molecular-weight 109289 #checksum 4583
Query Match	6.4%; Score 108; DB 2; Length 1001;
Best Local Similarity	33.3%; Pred. No. 1.23e-01;
Matches	20; Conservative 14; Mismatches 20; Indels 6; Gaps 6;
Db	97 FFAPTGVSTGVGNISPTT-FAGRMIMIAV-SVIGIPVNGILFAGL-GRYFG-RTFEATY 152
Qy	64 YFAFKICSGAA-NVVGPTMCFEDRMISPVKNVGRGLNALVNGTTGAVLGOKAPD-MY 121
RESULT	4
ENTRY	C69288
TITLE	#type complete hypothetical protein AF0307 - Archaeoglobus fulgidus
ORGANISM	#formal_name Archaeoglobus fulgidus

ENTRY	C75153	#type complete
TITLE	probable n2,n2-dimethylguanosine tRNA methyltransferase PAB2092 - Pyrococcus abyssi (strain Orsay)	
ORGANISM	#formal_name Pyrococcus abyssi	
DATE	20-Aug-1999	#sequence_revision 20-Aug-1999 #text_change
ACCESSIONS	C75153	
REFERENCE	A75001	
#authors	anonymous, Genoscope	
#submission	submitted to the EMBL Data Library, July 1999	
#description	Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution.	
#accession	C75153	
##status	preliminary	
##molecule_type	DNA	
##residues	1-383 ##label KAW	
##cross-references	GB:A7248284; GB:AL096836; NID:g5457730;	
##experimental_source	PIDN:CAB49306.1; PID:el515201; PID:g5457816	
GENETICS		
#gene	PAB2092	
SUMMARY	#length 383 #molecular-weight 43218 #checksum 9956	
Query Match	5.8%;	Score 97; DB 2; Length 383;
Best Local Similarity	21.6%;	Pred. No. 2.93e+00;
Matches	16;	Conservative 28; Mismatches 26; Indels 4; Gaps 4;
Db	308 QAVKLKILKDELDPVL-FYDTHG-LGRRLKIEARKI-EIINEL-RSLGYRASRTFSP	363
Qy	124 DVHLVKFLKEIPGALVILVSYDDPTGKMNDESKLFSDLGSSVAKQLGFRDSNVFICA	183
	: : : : : : : : : : : :	
Db	364 TGKTKDAPYEFVN	377
Qy	184 KDLRGKSPFEQFLK	197
	: : : : : : : : : : :	
RESULT	9	
ENTRY	G69269	#type complete
TITLE	molybdopterIn-binding oxidoreductase homolog - Archaeoglobus fulgidus	
ORGANISM	#formal_name Archaeoglobus fulgidus	
DATE	05-Dec-1997	#sequence_revision 05-Dec-1997 #text_change
ACCESSIONS	G69269	
REFERENCE	A69260	
#authors	Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.B.; Kervage, A.R.; Graham, D.E.; Kyriades, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.; Utterback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.P.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.	
#journal	Nature (1997)	390:364-370
#title	The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon Archaeoglobus fulgidus.	
#cross-references	MUID:98049343	
#accession	G69269	
##status	preliminary; nucleic acid sequence not shown; translation not shown	
##molecule_type	DNA	
##residues	1-1134 ##label KLE	
##cross-references	GB:AE001095; GB:AE000782; NID:g2689418; PID:g2650483; TIGR:AF0159	
SUMMARY	#length 1134 #molecular-weight 127688 #checksum 8700	
Query Match	5.8%;	Score 97; DB 2; Length 1134;
Best Local Similarity	27.3%;	Pred. No. 2.93e+00;

Matches 27; **Conservative** 23; **Mismatches** 40; **Indels** 9; **Gaps** 7;

Db VVAAEPTWVKGTIAIPYCMGRWADSAVKKPAYFRLNDGSAALYS--ELPDGA-SLPSS 1045
 ::::| |::||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:: ||:
Qy IMSPKVNNVGRC-LNIALVNVT-TGAVLGCKA-FDMYSGDVMMHLVKFLKEITPGGALLVA 144

```

Db 1046 DAYNPVKQLDETKKRILFTKSDRRYYDDLGI-DSWRFSG 1083
      :| :: :: ||: | :||: ||| |
Qy 145 SYDDPGTKMNDESRK-LFSDLGSSYAKQLGFRDSWVFIG 182

```

RESULT	10
ENTRY	T15631
TITLE	#type complete
ORGANISM	hypothetical protein C25p3.10 - <i>Caenorhabditis elegans</i>
DATE	#formal_name <i>Caenorhabditis elegans</i> 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

```

ACCESSIONS
REFERENCE
#authors
#submission
#description
#accession
#status
#molecule_type
#residues
#cross_references
#experimental_source

T15631
Z18379
Johnson, D.
submitted to the EMBL Data Library, June 1995
The sequence of C. elegans cosmid C25H3.
T15631
preliminary; translated from GB/EMBL/DDBJ
DNA
1-269
#label JOH
EMBL:U29535; NID:g868251; PID:g868259;
PIDN:AAA68788.1; CESP:C25H3.10
strain Bristol N2

```

```

GENEALICS
#gene CESP:C25H3.10
#introns 31/3; 68/3; 140/3; 188/3
#length 269 #molecular-weight 31750 #checksum 2802
SUMMARY

```

Query Match	5.7%;	Score 96;	DB 2;	Length 269;
Best Local Similarity	37.8%;	Pred. No. 3.87e+00;		
Matches	17;	Conservative	11;	Mismatches 14; Indels 3; Gaps 2;

```

Db-      148 NQESNFYHPLSGDHRK--FRGDWLEIQGNRLRS-SDISTETKN 189
          :|::|:: |:: |:: |:: |:: |:: |:: |:: |:: |::
QY       154 NDESRLKFLSDLGSSYAKQLGRDSWVFCAKDRLGKSPPEQFIKN 198

```

RESULT	11
ENTRY	Tl5849 #type complete
TITLE	hypothetical protein C56C10.9 - Caenorhabditis elegans
ORGANISM	#formal_name Caenorhabditis elegans
DATE	20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999

```

ACCESSIONS      C56C10.9
REFERENCE
#authors        Fulton, L.
#submission     submitted to the EMBL Data Library, June 1995
#description    The sequence of C. elegans cosmid C56C10.
#accession      T15849
#status         preliminary; translated from GB/EMBL/DBJ
#molecule_type DNA
#residues       1-320
##cross-references EMBL:U29488; NID:g868238; PID:g868247;
                  PID:AAA68777.1; CESP:C56C10.9
##experimental_source strain Bristol N2

```

```

#gene
#introns
SUMMARY
CESP:C56C10.9
73/2; 120/3; 150/1; 202/1; 279/3
#length 320 #molecular-weight 37556 #checksum 163

```

Query Match	5.7%;	Score 96;	DB 2;	Length 320;
Best Local Similarity	20.4%;	Pred. No. 3.87e+00;		
Matches	21;	Conservative	32;	Mismatches 43;
			Indels	7;
			Gaps	6;

```

Db      201 YDKNDEWSQDDFVNGIPGTVDLNPENME-KLEKERRLREFNEIDENS DGKATFR 259
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy      118 FDMYSGDVMHLVKFLKEIPGGALVLVASDDPGTKMNDSCR-KLFSD-LGSSVAKQLGFR 175

```

Db 260 ELYDVDPQFRLASKEVNDIMLTANDNDEKL-SLEELLERD 301
:
: : : : :
Yy 176 DSWYFIGAKDLR-GKSPFE-QFLKNSPTNYEGWPELLEME 215

RESULT	ENTRY	TITLE	ORGANISM	DATE
12	T00697	#type complete glycylpeptide N-tetradecanoyltransferase homolog F6E13.30 - Arabidopsis thaliana	#formal_name Arabidopsis thaliana #common_name mouse-ear cress	01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 30-Apr-1999

T00697
 #accessions
 Z14198
 #reference
 #authors
 Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby,
 M.L.; Brandon, R.C.; Sykes, S.M.; Mason, T.M.; Kerlavage,
 A.R.; Adams, M.D.; Somerville, C.R.; Venter, J.C.
 #submission
 submitted to the EMBL Data Library, June 1998
 #description
 Arabidopsis thaliana chromosome II BAC F6E13 genomic
 sequence.

```
#accession      T00697
##status        translated from GB/EMBL/DBDJ
##molecule_type DNA
##residues      1-389 #label ROU
##cross-references EMBL:AC004005; NID:g3212846; PID:g3212870
##experimental_source cultivar Columbia
```

```

GENEILCS
#map_position 2
#introns 100/3
#note FEEL3.30
SUMMARY #length 389 #molecular-weight 44450 #checksum 5324

```

Query Match	5.7%	Score 96;	DB 2;	Length 389;
Best Local Similarity	48.5%;	Pred. No.	3.87e+00;	
Matches	16;	Conservative	7;	Mismatches 8;
				Indels 2;
				Gaps 2;

```

Db 327 LVNDAL-IVSKQKGFDFVYASDVQMNESFLKEL 358
    ||| : | ||:| : |||: |||:
QY 104 LVNGTTGAVLGQKAFDM-YSGDVHMLVKFLKEI 135

```

```

RESULT 13
ENTRY
TITLE
B42804      #type complete
            immunodominant type-specific antigen tsW56 - Rickettsia
            tsutsugamushi
ORGANISM
            #formal_name Rickettsia tsutsugamushi
DATE
            01-Mar-1993 #sequence_revision 18-Nov-1994 #text_change
            20-Mar-1998
ACCESSIONS
B42804
REFERENCE
A42804
#authors
            Ohashi, N.; Nashimoto, H.; Ikeda, H.; Tamura, A.
#journal
            J. Biol. Chem. (1992) 267:12728-12735
#title
            Diversity of immunodominant 56-kDa type-specific antigen
            (TSA) of Rickettsia tsutsugamushi. Sequence and comparative
            analyses of the genes encoding TSA homologues from four
            antigenic variants.

```

```
#cross-references MUID:92316959
#accession B42804
#status preliminary
#molecule_type DNA
#residues 1-522 #label OHA
#cross-references GB:M63383; NID:gl52512; PID:gl52513
#experimental_source strain Kawasaki
#note sequence extracted from NCBI backbone (NCBIN:108848,
NCBIP:108849)
transmembrane protein
KEYWORDS
SUMMARY
#length 522 #molecular-weight 55971 #checksum 2488
```

Query Match	5.7%	Score 96;	DB 2;	Length 522;
Best Local Similarity	31.0%;	Pred. No. 3.87e+00;		
Matches	18;	Conservative	20;	Mismatches 17; Indels 3; Gaps 3;

[illegible]

W P E R L H (TM)

Release 3.1A John F. Collins, Biocomputing Research Unit.
Copyright (c) 1993-1998 University of Edinburgh, U.K.
Distribution rights by Oxford Molecular Ltd.

MPSrch_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Wed May 31 11:05:38 2000; MasPar time 6.14 Seconds
Tabular output not generated. 524.106 Million cell updates/sec

Title: >US-09-186-342-2
Description: (1-223) from US09186342.pep
Perfect Score: 1686
Sequence: 1 MRVSGVRLALIFAIVTW.....KYGWPELLEMEGMPKPKPF 223

Scoring table: PAM 150
Gap 11

Searched: 145341 seqs, 14437480 residues

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: a-issued
1:5A_COMB 2:5B_COMB 3:6_COMB 4:PCT_COMB 5:backfiles1

Statistics: Mean 30.975; Variance 139.082; scale 0.223

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query No.	Score	Match	Length	DB	ID	Description	Pred. No.
1	993	58.9	135	4	PCT-US95-0	Sequence 7, Applicatio	1.52e-80	
2	993	58.9	135	1	US-08-469-	Sequence 7, Applicatio	1.52e-80	
3	108	6.4	618	1	US-08-332-	Sequence 2, Applicatio	1.59e+00	
4	102	6.0	993	1	US-08-468-	Sequence 2, Applicatio	4.40e+00	
5	91	5.4	166	2	US-08-775-	Sequence 3, Applicatio	2.70e+01	
6	89	5.3	614	3	US-08-968-	Sequence 13, Applicati	3.72e+01	
7	89	5.3	680	1	US-07-674-	Sequence 2, Applicatio	3.72e+01	
8	89	5.3	680	2	US-08-436-	Sequence 2, Applicatio	3.72e+01	
9	86	5.1	731	1	US-08-731-	Sequence 2, Applicatio	6.01e+01	
10	85	5.0	76	1	US-08-478-	Sequence 9, Applicatio	7.03e+01	
11	85	5.0	76	1	US-07-881-	Sequence 9, Applicatio	7.03e+01	
12	85	5.0	76	1	US-08-120-	Sequence 9, Applicatio	7.03e+01	
13	84	5.0	119	4	PCT-US95-0	Sequence 8, Applicatio	8.23e+01	
14	84	5.0	119	1	US-08-249-	Sequence 8, Applicatio	8.23e+01	
15	84	5.0	119	2	US-08-886-	Sequence 8, Applicatio	8.23e+01	
16	84	5.0	433	1	US-08-700-	Sequence 20, Applicati	8.23e+01	
17	85	5.0	671	2	US-08-737-	Sequence 13, Applicati	7.03e+01	
18	83	4.9	191	1	US-07-885-	Sequence 29, Applicati	9.63e+01	
19	83	4.9	191	1	US-08-093-	Sequence 3, Applicatio	9.63e+01	
20	83	4.9	193	3	US-08-459-	Sequence 2, Applicatio	9.63e+01	
21	83	4.9	193	2	US-08-383-	Sequence 2, Applicatio	9.63e+01	
22	83	4.9	199	1	US-07-801-	Sequence 4, Applicatio	9.63e+01	
23	83	4.9	217	2	US-09-105-	Sequence 2, Applicatio	9.63e+01	

24	82	4.9	444	3	US-09-221-	Sequence 14, Applicati	1.13e+02
25	82	4.9	531	4	PCT-US92-0	Sequence 5, Applicatio	1.13e+02
26	82	4.9	533	4	PCT-US92-0	Sequence 3, Applicatio	1.13e+02
27	82	4.9	534	4	PCT-US92-0	Sequence 4, Applicatio	1.13e+02
28	82	4.9	627	2	US-08-700-	Sequence 6, Applicatio	1.13e+02
29	82	4.9	627	2	US-08-466-	Sequence 6, Applicatio	1.13e+02
30	82	4.9	627	3	US-08-467-	Sequence 6, Applicatio	1.13e+02
31	83	4.9	629	1	US-08-278-	Sequence 6, Applicatio	9.63e+01
32	83	4.9	629	3	US-08-464-	Sequence 6, Applicatio	9.63e+01
33	83	4.9	828	1	US-08-261-	Sequence 2, Applicatio	9.63e+01
34	83	4.9	2104	3	US-08-772-	Sequence 4, Applicatio	9.63e+01
35	83	4.9	2105	3	US-08-808-	Sequence 4, Applicatio	9.63e+01
36	83	4.9	2105	2	US-08-772-	Sequence 3, Applicatio	9.63e+01
37	83	4.9	2105	2	US-08-808-	Sequence 3, Applicatio	9.63e+01
38	81	4.8	611	2	US-08-677-	Sequence 2, Applicatio	1.31e+02
39	81	4.8	687	1	US-08-470-	Sequence 10, Applicati	1.31e+02
40	81	4.8	687	1	US-08-467-	Sequence 10, Applicati	1.31e+02
41	81	4.8	707	4	PCT-US95-1	Sequence 18, Applicati	1.31e+02
42	81	4.8	707	1	US-08-528-	Sequence 18, Applicati	1.31e+02
43	81	4.8	750	1	US-08-325-	Sequence 2, Applicatio	1.31e+02
44	81	4.8	750	2	US-08-394-	Sequence 2, Applicatio	1.31e+02
45	81	4.8	966	2	US-08-868-	Sequence 2, Applicatio	1.31e+02

ALIGNMENTS

RESULT 1
ID PCT-US95-07289-7 STANDARD; PRT; 135 AA.
XX
AC xxxxxx
DT
XX
DE Sequence 7, Application PC/TUS9507289
XX
XX
CC Sequence 7, Application PC/TUS9507289
CC GENERAL INFORMATION:
CC APPLICANT: Yu, Guo-Liang
CC APPLICANT: Rosen, Craig
CC TITLE OF INVENTION: Colon Specific Genes and Proteins
CC NUMBER OF SEQUENCES: 24
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
CC ADDRESSEE: Stewart & Olstein
CC STREET: 6 Becker Farm Road
CC CITY: Roseland
CC STATE: NJ
CC COUNTRY: USA
CC ZIP: 07068-1739
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: PCT/US95/07289
CC FILING DATE: 06-JUN-1995
CC CLASSIFICATION:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Ferraro, Gregory D.
CC REGISTRATION NUMBER: 36,134
CC REFERENCE/DOCKET NUMBER: 325800-265
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 201-994-1700
CC TELEFAX: 201-994-1744
CC INFORMATION FOR SEQ ID NO: 7:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 135 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 135 AA; 14852 MW; 99029 CN;


```
DE Sequence 2, Application US/08468557
XX
CC Sequence 2, Application US/08468557
CC Patent No. 5759833
CC GENERAL INFORMATION:
CC APPLICANT: Shiba, Kiyotaka
CC APPLICANT: Kranz, Janice E.
CC APPLICANT: Schimmel, Paul R.
CC TITLE OF INVENTION: Human Isoleucyl-tRNA Synthetase
CC TITLE OF INVENTION: Proteins, Nucleic Acids and Tester Strains Comprising Same
CC NUMBER OF SEQUENCES: 26
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P. C.
CC STREET: Two Militia Drive
CC CITY: Lexington
CC STATE: Massachusetts
CC COUNTRY: US
CC ZIP: 02173
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: PatentIn Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/468,557
CC FILING DATE: 06-JUN-1995
CC CLASSIFICATION: 435
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER: US 08/250,852
CC FILING DATE: 27-MAY-1994
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Brook, David E.
CC REGISTRATION NUMBER: 22,592
CC REFERENCE/DOCKET NUMBER: CPI94-13A
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 617-861-6240
CC TELEFAX: 617-861-9540
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 993 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 993 AA; 111800 MW; 5344624 CN;

Query Match 6.08; Score 102; DB 1; Length 993;
Best Local Similarity 30.1%; Pred. No. 4.40e+00;
Matches 22; Conservative 23; Mismatches 23; Indels 5; Gaps 5;

Db 887 VIEPGLLFEIEMLOSSTSQ-LN-ELMMASESTLLAQEPREM-TADVIELKGFLLI 943
QY 76 VVGPTMCFEDR-MIMSPVKNNVGRGLNIALYNGTTGAVLGQKAFDMYSGDVMHL-VKFLK 133

Db 944 NLEGGDIRRESSY 956
QY 134 EIPGALVLVASY 146

RESULT 5
ID US-08-775-978-3 STANDARD; PRT; 166 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 3, Application US/08775978
CC Sequence 3, Application US/08775978
CC Patent No. 5831049
CC GENERAL INFORMATION:
CC APPLICANT: HILLMAN, JENNIFER L.
CC APPLICANT: GOLI, SURYA K.
CC TITLE OF INVENTION: NOVEL HUMAN THIOREDOXIN
```

```
CC NUMBER OF SEQUENCES: 3
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Incyte Pharmaceuticals, Inc.
CC STREET: 3174 Porter Drive
CC CITY: Palo Alto
CC STATE: CA
CC COUNTRY: USA
CC ZIP: 94304
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Diskette
CC COMPUTER: IBM Compatible
CC OPERATING SYSTEM: DOS
CC SOFTWARE: FastSEQ for Windows Version 2.0
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/775,978
CC FILING DATE: To Be Assigned
CC PRIOR APPLICATION DATA:
CC APPLICATION NUMBER:
CC FILING DATE:
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Billings, Lucy J.
CC REGISTRATION NUMBER: 36,749
CC REFERENCE/DOCKET NUMBER: PF-0176 US
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 415-855-0555
CC TELEFAX: 415-845-4166
CC INFORMATION FOR SEQ ID NO: 3:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 166 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC IMMEDIATE SOURCE:
CC LIBRARY: GenBank
CC CLONE: 1545817
CC SEQUENCE 166 AA; 18416 MW; 148927 CN;

Query Match 5.4%; Score 91; DB 2; Length 166;
Best Local Similarity 19.2%; Pred. No. 2.70e+01;
Matches 15; Conservative 27; Mismatches 31; Indels 5; Gaps 5;

Db 1 MAORLLLRRLTSIIISGKPSQSRWAPVASRALKTPQYSPGYLTVPQARSYITTRVCGST 60
QY 17 VVTWTFIRSYMSFSMKTIR-LPRWLASPTKEIQVKYKCG-L-IKPCPA-NVPAFKICSG 72

Db 61 TENIQDGPD-FQDRVVNS 77
QY 73 AANVVGPTMCFEDRMIMS 90

RESULT 6
ID US-08-968-563-13 STANDARD; PRT; 614 AA.
XX
AC xxxxxx
XX
DT
XX
DE Sequence 13, Application US/08968563
CC Sequence 13, Application US/08968563
CC Patent No. 6013494
CC GENERAL INFORMATION:
CC APPLICANT: CHARLES E. NAKAMURA
CC APPLICANT: ANTHONY A. GATENBY
CC APPLICANT: AMY (KUANG-HUA) HSU
CC APPLICANT: RICHARD D. LA REAU
CC APPLICANT: SHARON L. HAYNIE
CC APPLICANT: MARIA DIAZ-TORRES
CC APPLICANT: DONALD E. TRIMBUR
CC APPLICANT: GREGORY M. WHITED
CC APPLICANT: VASANTHA NAGARAJAN
CC APPLICANT: MARK S. PAYNE
CC APPLICANT: STEPHEN K. PICATAGGIO
```


CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/436,900A
CC FILING DATE: 08-MAY-1995
CC CLASSIFICATION: 536
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Barnhard, Elizabeth M.
CC REGISTRATION NUMBER: 31,088
CC REFERENCE/DOCKET NUMBER: 31,104-03
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 201-683-2158
CC TELEFAX: 201-683-4117
CC INFORMATION FOR SEQ ID NO: 2:
CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 680 amino acids
CC TYPE: amino acid
CC STRANDEDNESS: single
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 680 AA; 73675 MW; 2557540 CN;

Query Match 5.3%; Score 89; DB 2; Length 680;
Best Local Similarity 28.1%; Pred. No. 3.72e+01;
Matches 16; Conservative 20; Mismatches 17; Indels 4; Gaps 4;

Db 81 VSETIRKGLIDVEMYNSTOGLMAGSVSAMFGSAVNQVLVAFK-LPISGTHCIVGA 136
QY 92 VKNNVGRGL-NIALVNGTGVAVLGQKAFDMYSGDVNHLV-KFLKEIP-GGALVVLVAS 145

RESULT 9
ID US-08-731-716-2 STANDARD; PRT; 731 AA.
XX
AC xxxxxx
XX
DT
DE
DE 2, Application US/08731716
XX
XX Sequence 2, Application US/08731716
CC Patent No. 5789202
CC GENERAL INFORMATION:
CC APPLICANT: Hoskins, JoAnn
CC APPLICANT: Jaskunas, S. Richard
CC APPLICANT: Rocky, Pamela K.
CC APPLICANT: Zhao, Genshi
CC APPLICANT: Rosteck, Paul R. Jr.
CC APPLICANT: No. 5789202ris, Franklin H.
CC TITLE OF INVENTION: Penicillin Binding Protein From
CC TITLE OF INVENTION: Streptococcus Pneumoniae
CC NUMBER OF SEQUENCES: 3
CC CORRESPONDENCE ADDRESS:
CC ADDRESSEE: Eli Lilly and Company
CC STREET: Lilly Corporate Center
CC CITY: Indianapolis
CC STATE: Indiana
CC COUNTRY: U.S.
CC ZIP: 46285
CC COMPUTER READABLE FORM:
CC MEDIUM TYPE: Floppy disk
CC COMPUTER: IBM PC compatible
CC OPERATING SYSTEM: PC-DOS/MS-DOS
CC SOFTWARE: Patent In Release #1.0, Version #1.30
CC CURRENT APPLICATION DATA:
CC APPLICATION NUMBER: US/08/731,716
CC FILING DATE:
CC CLASSIFICATION: 435
CC ATTORNEY/AGENT INFORMATION:
CC NAME: Webster, Thomas D.
CC REGISTRATION NUMBER: 39,872
CC REFERENCE/DOCKET NUMBER: X-10,887
CC TELECOMMUNICATION INFORMATION:
CC TELEPHONE: 317-276-3334
CC INFORMATION FOR SEQ ID NO: 2:

CC SEQUENCE CHARACTERISTICS:
CC LENGTH: 731 amino acids
CC TYPE: amino acid
CC TOPOLOGY: linear
CC MOLECULE TYPE: protein
CC SEQUENCE 731 AA; 80798 MW; 2731494 CN;

Query Match 5.1%; Score 86; DB 1; Length 731;
Best Local Similarity 21.6%; Pred. No. 6.01e+01;
Matches 19; Conservative 24; Mismatches 43; Indels 2; Gaps 2;

Db 472 ATVNDLVGDKAFKFEKGLMEKVRVLGVAGLVSGVETNPQLMAQAYAFANGLMPEA 531
QY 72 GAANVVGPTCMCFEDMIMSPVKNNVGRGLNIALVNGTGVAVLG-QKAFDMYSGD-VMLHV 129
Db 532 HFISRIENASQVVIASHKNSQKRVIDS 559
QY 130 KFLKEIPGGALVVLVASYDDPGTKMNDES 157

RESULT 10
ID US-08-478-675-9 STANDARD; PRT; 76 AA.
XX
AC xxxxxx

Sequence 9, Application US/08478675

Sequence 9, Application US/08478675
Patent No. 5773246
GENERAL INFORMATION:
APPLICANT: KEENE, JACK D.
APPLICANT: KING, PETER H.
APPLICANT: LEVINE, TODD
TITLE OF INVENTION: METHODS AND COMPOSITIONS USEFUL IN THE
TITLE OF INVENTION: RECOGNITION, BINDING AND EXPRESSION OF RIBONUCLEIC ACID
TITLE OF INVENTION: INVOLVED IN CELL GROWTH, NEOPLASIA AND IMMUNOREGULATION
NUMBER OF SEQUENCES: 101
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,675
FILING DATE: 07-JUN-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/120,827
FILING DATE: 15-SEP-1993
ATTORNEY/AGENT INFORMATION:
NAME: Oblon, No. 5773246man F.
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 714-158-0 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)413-3000
TELEFAX: (703)413-2220
TELEX: 248655 OPAT UR
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 76 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: peptide

ID PCT-US95-06764-8 STANDARD; PRT; 119 AA.
XX
AC xxxxxx

Sequence 8, Application PC/TUS9506764

Sequence 8, Application PC/TUS9506764

GENERAL INFORMATION:

APPLICANT: The Regents of the University of California

TITLE OF INVENTION: CLONED leptospira OUTER MEMBRANE PROTEIN

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson

STREET: 4225 Executive Square, Suite 1400

CITY: La Jolla

STATE: California

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/06764

FILING DATE: 25-MAY-1995

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Haile, Ph.D., Lisa A.,

REGISTRATION NUMBER: 38,347

REFERENCE/DOCKET NUMBER: FD3602

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 678-5070

TELEFAX: (619) 678-5099

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 119 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

IMMEDIATE SOURCE:

CLONE: Pupa

FEATURE:

NAME/KEY: Protein

LOCATION: 1

SEQUENCE 119 AA; 12942 MW; 73190 CN;

Query Match 5.0%; Score 84; DB 4; Length 119;

Best Local Similarity 29.3%; Pred. No. 8.23e+01;

Matches 22; Conservative 21; Mismatches 27; Indels 5; Gaps 4;

Db 47 EIVRGATGLMTGAGDPSAVVN-VIRKLDPEVGKNY--ELGWKDGAEKGVDAITLNNNIF 103

Qy 134 EIPGGALVLVASYDDPGTKMNDESRKLFSDLGSSYAKQLGFRDSWVFICAKD-LRGSPPF 192

Db 104 DKYPNATVTLRYD 118

Qy 193 EQ-FLKNSPDTNKYE 206

RESULT 14

ID US-08-249-013-8 STANDARD; PRT; 119 AA.

XX

AC xxxxxx

XX

DT

DE

XX

XX

XX

CC

Sequence 8, Application US/08249013

Sequence 8, Application US/08249013

CC Patent No. 5643754

CC GENERAL INFORMATION:

CC APPLICANT: Haake, David A.

CC TITLE OF INVENTION: CLONED leptospira OUTER MEMBRANE PROTEIN

CC NUMBER OF SEQUENCES: 10

CC CORRESPONDENCE ADDRESS:

CC ADDRESSEE: Spensley Horn Jubas & Lubitz

CC STREET: 1880 Century Park East, Suite 500

CC CITY: Los Angeles

CC STATE: California

CC COUNTRY: USA

CC ZIP: 90067

CC COMPUTER READABLE FORM:

CC MEDIUM TYPE: Floppy disk

CC COMPUTER: IBM PC compatible

CC OPERATING SYSTEM: PC-DOS/MS-DOS

CC SOFTWARE: Patent In Release #1.0, Version #1.25

CC CURRENT APPLICATION DATA:

CC APPLICATION NUMBER: US/08/249,013

CC FILING DATE: 25-MAY-1994

CC CLASSIFICATION: 435

CC ATTORNEY/AGENT INFORMATION:

CC NAME: Tumarkin Ph.D., Lisa A.,

CC REGISTRATION NUMBER: P-38,347

CC REFERENCE/DOCKET NUMBER: PD-3602

CC TELECOMMUNICATION INFORMATION:

CC TELEPHONE: (619) 455-5100

CC TELEFAX: (619) 455-5110

CC INFORMATION FOR SEQ ID NO: 8:

CC SEQUENCE CHARACTERISTICS:

CC LENGTH: 119 amino acids

CC TYPE: amino acid

CC STRANDEDNESS: single

CC TOPOLOGY: linear

CC MOLECULE TYPE: protein

CC IMMEDIATE SOURCE:

CC CLONE: Pupa

CC FEATURE:

CC NAME/KEY: Protein

CC LOCATION: 1

CC SEQUENCE 119 AA; 12942 MW; 73190 CN;

Query Match 5.0%; Score 84; DB 1; Length 119;

Best Local Similarity 29.3%; Pred. No. 8.23e+01;

Matches 22; Conservative 21; Mismatches 27; Indels 5; Gaps 4;

Db 47 EIVRGATGLMTGAGDPSAVVN-VIRKLDPEVGKNY--ELGWKDGAEKGVDAITLNNNIF 103

Qy 134 EIPGGALVLVASYDDPGTKMNDESRKLFSDLGSSYAKQLGFRDSWVFICAKD-LRGSPPF 192

Db 104 DKYPNATVTLRYD 118

Qy 193 EQ-FLKNSPDTNKYE 206

RESULT 15

ID US-08-886-863-8 STANDARD; PRT; 119 AA.

XX

AC xxxxxx

XX

DT

XX

DE

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Sequence 8, Application US/08886863

Sequence 8, Application US/08886863

Patent No. 5824321

GENERAL INFORMATION:

APPLICANT: Haake, David A.

TITLE OF INVENTION: CLONED leptospira OUTER MEMBRANE PROTEIN

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Spensley Horn Jubas & Lubitz

STREET: 1880 Century Park East, Suite 500

CC CITY: Los Angeles
 CC STATE: California
 CC COUNTRY: USA
 CC ZIP: 90067
 CC COMPUTER READABLE FORM:
 CC MEDIUM TYPE: Floppy disk
 CC COMPUTER: IBM PC compatible
 CC OPERATING SYSTEM: PC-DOS/MS-DOS
 CC SOFTWARE: Patent in Release #1.0, Version #1.25
 CC CURRENT APPLICATION DATA:
 CC APPLICATION NUMBER: US/08/886.863
 CC FILING DATE: 01-JUL-1997
 CC CLASSIFICATION: 435
 CC PRIOR APPLICATION DATA:
 CC APPLICATION NUMBER: US 08/249,013
 CC FILING DATE: 25-MAY-1994
 CC ATTORNEY/AGENT INFORMATION:
 CC NAME: Tumarkin Ph.D., Lisa A.,
 CC REGISTRATION NUMBER: P-38,347
 CC REFERENCE/DOCKET NUMBER: PD-3602
 CC TELECOMMUNICATION INFORMATION:
 CC TELEPHONE: (619) 455-5100
 CC TELEFAX: (619) 455-5110
 CC INFORMATION FOR SEQ ID NO: 8:
 CC SEQUENCE CHARACTERISTICS:
 CC LENGTH: 119 amino acids
 CC TYPE: amino acid
 CC STRANDEDNESS: single
 CC TOPOLOGY: linear
 CC MOLECULE TYPE: protein
 CC IMMEDIATE SOURCE:
 CC CLONE: Pupa
 CC FEATURE:
 CC NAME/KEY: Protein
 CC LOCATION: 1
 CC
 SQ SEQUENCE 119 AA; 12942 MW; 73190 CN;

Query Match 5.08; Score 84; DB 2; Length 119;
 Best Local Similarity 29.38; Pred. No. 8.23e+01;
 Matches 22; Conservative 21; Mismatches 27; Indels 5; Gaps 4;
 Db 47 EIVRGATGLMTGAGDPSAVVN-VIRKLDPEVGKYN--ELGHWKDGAEKGVDTATLNVNNIF 103
 QY 134 EIPGALVLVASYDDPGTKMNDSEKRLFLSDLGSSYAKQLGDFRDSWVFIGAKD-LRGKSPF 192
 Db 104 DKYPRNATVTLRYD 118
 QY 193 EQ-FLKNSPDTNKYE 206

Search completed: Wed May 31 11:05:48 2000
 Job time : 10 secs.

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 31, 2000, 17:51:15 ; Search time 799,5 Seconds
(without alignments)
3391.628 Million cell updates/sec

Title: US-09-186-342-8
Perfect score: 669
Sequence: 1 atgmngntnwsngngngtnyt.....gyatgccnccnaarccntty 669

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :	EST:*
	1: em_est1:*
	2: em_est2:*
	3: em_est3:*
	4: em_est4:*
	5: em_est5:*
	6: em_est6:*
	7: em_est7:*
	8: em_est8:*
	9: em_est9:*
	10: em_est10:*
	11: em_est11:*
	12: em_est12:*
	13: em_est13:*
	14: em_est14:*
	15: em_est15:*
	16: em_est16:*
	17: em_est17:*
	18: em_est18:*
	19: em_est19:*
	20: gb_est1:*
	21: gb_est2:*
	22: gb_est3:*
	23: gb_est4:*
	24: gb_est5:*
	25: gb_est6:*
	26: gb_est7:*
	27: gb_est8:*
	28: gb_est9:*
	29: gb_est10:*
	30: gb_est11:*
	31: gb_est12:*
	32: gb_est13:*
	33: gb_est14:*
	34: gb_est15:*
	35: gb_est16:*
	36: gb_est17:*
	37: gb_est18:*
	38: gb_est19:*
	39: gb_est20:*
	40: gb_est21:*
	41: gb_est22:*
	42: gb_est23:*
	43: gb_est24:*
	44: gb_est25:*
	45: gb_est26:*
	46: gb_est27:*
	47: gb_est28:*
	48: gb_est29:*
	49: gb_est30:*
	50: gb_est31:*
	51: gb_est32:*
	52: em_est20:*
	53: em_est21:*
	54: em_est22:*
	55: em_est23:*
	56: em_est24:*
	57: em_est25:*
	58: em_est26:*
	59: gb_est33:*
	60: gb_est34:*
	61: gb_est35:*
	62: gb_est36:*
	63: gb_est37:*
	64: gb_est38:*
	65: em_est27:*
	66: em_est28:*
	67: em_est29:*
	68: em_est30:*
	69: gb_est39:*
	70: gb_est40:*
	71: gb_est41:*
	72: gb_est42:*
	73: gb_est43:*
	74: gb_est44:*
	75: em_est31:*
	76: em_est32:*
	77: em_est33:*
	78: em_est34:*
	79: gb_est45:*
	80: gb_est46:*
	81: gb_est47:*
	82: gb_gss1:*
	83: gb_gss2:*
	84: gb_gss3:*
	85: gb_gss4:*
	86: em_gss1:*
	87: em_gss2:*
	88: em_gss3:*
	89: em_gss4:*
	90: gb_gss5:*
	91: gb_gss6:*
	92: gb_gss7:*
	93: gb_gss8:*
	94: gb_gss9:*
	95: em_gss5:*
	96: em_gss6:*
	97: em_gss7:*
	98: em_gss8:*
	99: em_gss9:*
	100: em_gss10:*
	101: em_gss11:*
	102: gb_gss10:*
	103: gb_gss11:*
	104: em_gss12:*
	105: gb_gss12:*
	106: gb_gss13:*
	107: gb_gss14:*
	108: gb_gss15:*
	109: gb_gss16:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	%	Query
--------	---	-------

No.	Score	Match	Length	DB	ID	Description
C 1	224	33.5	645	79	AW300770	AW300770 xk05g02.x
C 2	223.8	33.5	668	63	AW001287	AW001287 wu27e06.x
C 3	216.6	32.4	306	31	AA297147	AA297147 EST112731
C 4	212.6	31.8	778	62	AI983327	AI983327 fc6e001.y
C 5	204	30.5	525	80	AW323823	AW323823 uo77a04.y
C 6	203.2	30.4	615	50	AI691072	AI691072 wf22c03.x
C 7	200.8	30.0	646	62	AI936111	AI936111 wo61h03.x
C 8	192.8	28.8	461	36	AA611725	AA611725 vo89f11.r
C 9	189.8	28.4	468	31	AA297176	AA297176 EST112740
C 10	187.6	28.0	417	31	AA298484	AA298484 EST114112
C 11	173.2	25.9	387	33	AA451058	AA451058 vf86b07.r
C 12	161.4	24.1	518	35	AA563933	AA563933 nk23b02.s
C 13	157.6	23.6	410	24	NA1338	NA1338 yw68g05.rl
C 14	157	23.5	570	74	AW175415	AW175415 f135h03.y
C 15	154.6	23.1	376	23	H26710	H26710 y114b12.r1
C 16	150.6	22.5	564	36	AA622758	AA622758 np76c04.s
C 17	149	22.3	550	81	AW455134	AW455134 zeh12767
C 18	144.2	21.6	264	38	AY793766	AY793766 vr96e04.r
C 19	144	21.5	538	44	AI245950	AI245950 qk45b04.x
C 20	135.8	20.3	364	80	AW369250	AW369250 QV1-BN000
C 21	126.6	18.9	498	45	AI348065	AI348065 qp56d09.x
C 22	124.6	18.6	519	35	AA592348	AA592348 vo27a08.r
C 23	124	18.5	464	39	AA847242	AA847242 of01b11.s
C 24	122.4	18.3	379	80	AW315283	AW315283 12472.MAR
C 25	122	18.2	518	36	AA622120	AA622120 nq56b03.s
C 26	121.6	18.2	508	60	AT801582	AT801582 to91c11.x
C 27	120.4	18.0	461	36	AA622370	AA622370 np21f04.s
C 28	119.4	17.8	459	63	AW001308	AW001308 wu28c10.x
C 29	119.4	17.8	550	64	AW045110	AW045110 um17g12.y
C 30	118.8	17.8	594	35	AA552519	AA552519 nk14e01.s
C 31	118.4	17.7	232	27	AA047528	AA047528 zk73g01.r
C 32	118	17.6	542	79	AA258184	AA258184 uq31e05.y
C 33	117.6	17.6	338	32	AA362591	AA362591 EST72523
C 34	117.6	17.6	448	26	W54153	W54153 md17f04.r1
C 35	111.8	16.7	286	63	AW107522	AW107522 um17g12.x
C 36	109.8	16.4	485	69	AI964291	AI964291 EST269405
C 37	108.4	16.2	244	32	AA379625	AA379625 EST92751
C 38	108	16.1	361	33	AA408987	AA408987 EST03631
C 39	108	16.1	509	34	AA512284	AA512284 vj19g03.r
C 40	103.6	15.5	248	28	AA078382	AA078382 7P02F05.C
C 41	101	15.1	561	34	AA475410	AA475410 vhl7h03.r
C 42	100.6	15.0	710	41	AT038272	AT038272 oy85g06.x
C 43	99.8	14.9	311	80	AW315346	AW315346 12597.MAR
C 44	94	14.1	573	33	AA402132	AA402132 zu55b03.r
C 45	92.6	13.8	367	31	AA282522	AA282522 PMY0756.K

ALIGNMENTS

```

RESULT      1
AW300770/c
LOCUS
DEFINITION
    similar to SWI_CGAP_Col9 Homo sapiens cDNA clone IMAGE:2665874 3'
    xk0502_1 NCIEF07_MOUSE P97805 PROTEIN EF-7 ;, mRNA sequence.
ACCESSION  AW300770
VERSION     AW300770.1
KEYWORDS   AW300770.1  GI:6710447
SOURCE     EST.
            human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 645)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE      National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL    Unpublished (1997)
COMMENT    On Jul 8, 1999 this sequence version replaced gi:5420873.
            Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov

```


similar to SW:EF07_MOUSE P97805 PROTEIN EF-7 ;, mRNA sequence.
 AI936111.1 GI:5674981
 EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 646)
 REFERENCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 TITLE Unpublished (1997)
 JOURNAL On May 18, 1998 this sequence version replaced gi:3138334.
 COMMENT Contact: Robert Strausberg, Ph.D.
 Tel: (301) 496-1550
 Email: Robert.Strausberg@nih.gov
 Tissue Procurement: Michael J. Brownstein, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/dbbrp/image/image.html
 Seq primer: -40UP from Gibco
 High quality sequence stop: 438.
 Location/Qualifiers
 1..646
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:2459861"
 /clone_lib="NCI_CGAP_Pr22"
 /sex="male"
 /tissue_type="normal prostate"
 /lab_host="DH10B"
 /note="Organ: prostate; Vector: pT73D-Pac (Pharmacia)
 with a modified polylinker; 1st strand cDNA was prepared
 from normal prostate bulk tissue, and was then primed with
 a Not I - oligo(dT) primer. Double-stranded cDNA was
 ligated to Eco RI adaptors (Pharmacia), digested with Not
 I and cloned into the Not I and Eco RI sites of the
 modified pT73 vector. Library is normalized, and was
 constructed by Bento Soares and M. Fatima Bonaldo."
 134 a 188 c 163 g 161 t
 BASE COUNT
 ORIGIN
 Query Match 30.0%; Score 200.8; DB 62; Length 646;
 Best Local Similarity 53.2%; Pred. NO. 1.4e-42;
 Matches 167; Conservative 76; Mismatches 71; Indels 0; Gaps 0;
 QY 356 avatgataywngngnagatnatgcaaytngtngaarttyttnaargarathccngngng 415
 Db 645 ACATGTACTACCGAATGTATGCTAGTGAAGTCTCTAAAGAAAGTGGTGGG 586
 QY 416 cnyntngtngtngcnwsntaygaygcyccngngnacnaaratgaaygaygawsmgna 475
 Db 585 CACTGGTGTGGAGGCTTCATCCCGCGAACCAGGACCTAAATGATCGATGAAGCAGTA 526
 QY 476 arytttysngagytngngnwsntaygcnaarcararytngntngngaywsntggg 535
 Db 525 AACTCTTCTGACTGCTGGGAGTTTCATACGATACAACTGGCTTCGGGACAGCTGGG 466
 QY 536 tnttathggngcnaragaytngmngngnaarwsnccnttygarcarttyttnaaraayw 595
 Db 465 TCTTCATATGACCCAGACACCTCAGGGGTAAAGCCCCCTTTGAGCAGTCTTAAGAACA 406
 QY 596 snccngayacnaaayartaygargntggccngarytyntngnatgatgargntggtatgc 655
 Db 405 GCCCAGACACATCATATACGAGGGATGCCAGAGCTGCTCGAGATGGAGGGCTGCATGC 346

QY 656 cncnnaareccntty 669
 Db 345 CCCGGAAGCCATTT 332
 RESULT 8
 AA611725 461 bp mRNA EST 01-OCT-1997
 LOCUS AA611725
 DEFINITION vo89f11.r1 Barstead mouse irradiated colon MRLR87 Mus musculus cDNA
 clone IMAGE:1066317 5' similar to TR:G452270 G452270 2-19 PROTEIN
 PRECURSOR. ;, mRNA sequence.
 ACCESSION AA611725
 VERSION AA611725.1 GI:2461760
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 461)
 REFERENCE
 AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
 Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
 Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
 Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
 Waterston,R.
 TITLE The WashU-HHMI Mouse EST Project
 JOURNAL Unpublished (1996)
 COMMENT On Sep 12, 1996 this sequence version replaced gi:1405306.
 Contact: Marra M/Mouse EST Project
 WashU-HHMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: mouseest@watson.wustl.edu
 This clone is available royalty-free through LLNL; contact the
 IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:588677
 Seq primer: -28ml3 rev2 ET from Amersham
 High quality sequence stop: 422.
 Location/Qualifiers
 1..461
 /organism="Mus musculus"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:1066317"
 /clone_lib="Barstead mouse irradiated colon MRLR87"
 /dev_stage="8 weeks"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (Pharmacia) with a modified
 polylinker; Site 1: EcoRI; Site 2: NotI; Tissue obtained
 from 8 week old mouse. Colon was harvested 72 hours after
 irradiation with 1400 Gys. 1st strand cDNA was primed
 with a Not I - oligo(dT) primer
 [5'TGTAGCAATCTGAAGTGGGAGCGCCGCTTTTCTTTTCTTTTCTTTT
 T 3']; double-stranded cDNA was ligated to Eco RI
 adaptors [AATTCGATCCTTG], digested with Not I and cloned
 into the Not I and Eco RI sites of the modified pT73
 vector. Library constructed by Bob Barstead."
 136 a 108 c 127 g 89 t 1 others
 BASE COUNT
 ORIGIN
 Query Match 28.8%; Score 192.8; DB 36; Length 461;
 Best Local Similarity 49.6%; Pred. NO. 1.5e-40;
 Matches 172; Conservative 75; Mismatches 100; Indels 0; Gaps 0;
 QY 317 ayggnacnacngngcngtngtngnccaraargccttygayatgtaywngngaygtna 376
 Db 15 ACGGATCAACTGCCAGGTGATGAAAGAGCAGACAGCTTGACATGTACTCTGGAGACCC 74
 QY 377 tgcaytngttnaarttyttnaargarathccngngngcnyntngtngtngtngcnwsnt 436
 Db 75 AACTCCTGCTGAACTCTTACGGNAATCCAGATAGACCCCTGGTGGTAGCCTCCT 134

For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (<http://www.tigr.org/tdb/hgi/hgi.html>)
seq primer: M13 Reverse.

```

FEATURES
  source
    Location/Qualifiers
      1..417
        /organism="Homo sapiens"
        /db_xref="ATCC (inhost):175003"
        /db_xref="taxon:9606"
        /clone_lib="Prostate gland I"
        /sex="male"
        /dev_stage="adult, 21 yrs"
        /note="Organ: prostate; Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"
BASE COUNT      100 a      114 c      106 g      90 t      7 others
ORIGIN
Query Match      28.0%; Score 187.6; DB 31; Length 417;
Best Local Similarity 62.6%; Pred. No. 3.3e-39;
Matches 181; Conservative 56; Mismatches 50; Indels 2; Gaps 2;

```

[illegible]

RESULT	11
AA451058	
LOCUS	
DEFINITION	IMAGE:07017.1 Soares_mammary_gland_NbMMG Mus musculus cDNA clone IMAGE:850717.5', similar to TR:G452270 G452270 2-19 PROTEIN PRECURSOR. ;, mRNA sequence.
ACCESSION	AA451058
VERSION	AA451058.1
KEYWORDS	GI:2164728
SOURCE	EST.
ORGANISM	house mouse.
	Mus musculus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE	1 (bases 1 to 387)
AUTHORS	Marra,M., Hliller,D., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le.M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.

TITLE The WashU-HHMI Mouse EST Project
JOURNAL Unpublished (1996)
COMMENT On Sep 12, 1996 this sequence version replaced gi:1394413.
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of MedicineP
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu
This clone is available royalty-free through LLNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information
 MG1:502869
 Seq primer: -28m13 rev2 ET from Amersham
 High quality sequence stop: 382.

```
FEATURES
source
    i. .387
    Location/Qualifiers
        magn_qualifier Sequence_SeqP: 302.
        /organism="Mus musculus"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone_image="IMAGE:850717"
        /clone_lib="Soares_mammary_gland_NDMMG"
        /sex="male"
        /tissue_type="mammary gland"
        /dev_stage="4 weeks"
        /lab_host="DH10A"
```

BASE COUNT 109 a 95 q 82 t

```
Query Match      25.9%; Score 173.2; DB 33; Length 387;
Best Local Similarity 49.5%; Pred. No. 2.1e-35;
Matches 150; Conservative 71; Mismatches 82; Indels 0; Gaps 0;
```

[illegible][illegible]

RESULT	12
AA563933/c	
LOCUS	AA563933
	518 bp mRNA EST
DEFINITION	nk23b02.s1 NCI_CGAP_Coll Homo sapiens cDNA clone IMAGE:1014315 3'
	similar to TR:G452270 G452270 2-19 PROTEIN PRECURSOR. ; mRNA

AA563933
AA563933.1
EST.
human.

ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 518)
REFERENCE	


```

Qy 248 tygargaymgnatgathatqwsnccngntnaaraayaaygtnggmgnngnytnaayathg 307
Db 3 TCGAGGACACAGATGCTGATGAGCAGCTCAAGGACACACGTGGGCCCGGNTGAAACATCG 62

Qy 308 cnytngtnaayggnacnacngngcngntnytingnncaraargcnytygayatgtaywsng 367
Db 63 CCCTGGTGAACGGGGTCAAGCGCGG-ANTCATCGAGGCCCGGGCTTTTGACATGTNGGCCG 121

Qy 368 gngaaygtnatgcayyytngtnaarttytnaargarathccngngngcnytngtnyng 427
Db 122 GAGATGTCAACGACCTTTGAAGTTTATTCGGCCACTGCACGAAGGCACCCCTNGTGTTCG 181

Qy 428 tngcnwsntaygayayccnggnacnaaratgaaygaygarwsnmgnnaarytnttywsng 487
Db 182 TGGCATCTAGGAGCACCACCAAGAGATGAATGAGAGACCAAGAAAGCTCTTCAGTG 241

Qy 488 ayytnggnwnwsntaygcnaarcarxytngnttymgngaywsntgggtnttyathggng 547
Db 242 AGCTGGGCAGCAGGACGCCCAAGGAGCTGGCTTCCGGGACAC-TNGGTGTTTTTCGGGG 300

Qy 548 cnaargayytmngngnaarwsnccnttygarcarttytnaaraaywsnccngayac 605
Db 301 CCAAGGGTGTNCAGAACAGAGGCCCTTTNAGCAGCACGTGAAGACAGTTAAGGCAC 358

```

Search completed: May 31, 2000, 17:51:19
Job time: 3436 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 31, 2000, 19:14:15 ; Search time 57.43 Seconds
(without alignments)
1514.191 Million cell updates/sec

Title: US-09-186-342-8
Perfect score: 669
Sequence: 1 atgmgngtnwngngngtnt.....gyatgcnccnaarcontty 669

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 230463 seqs, 64992525 residues

Total number of hits satisfying chosen parameters: 460926

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/5C_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/5D_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/6_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
7: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	294.6	44.0	548	2	US-08-469-667-6
2	294.6	44.0	548	6	PCT-US95-07289-6
C 3	48.6	7.3	7218	1	US-08-232-463-14
C 4	39.4	5.9	1215	4	US-09-092-770-8
5	35.8	5.4	1058	2	US-08-608-241-3
6	35.8	5.4	1058	3	US-08-922-182-3
7	35.8	5.4	1058	3	US-08-919-953-3
8	35.4	5.3	2262	4	US-08-674-887A-5
C 9	35.2	5.3	774	6	PCT-US91-05766-1
10	35.2	5.3	1564	4	US-08-852-481-1
11	35.2	5.3	3101	4	US-08-868-786-1
C 12	35	5.2	2855	4	US-08-852-153-1
C 13	35	5.2	3255	4	US-08-852-153-5
C 14	35	5.2	3327	4	US-08-852-153-3
15	34.4	5.1	579	2	US-08-190-560-3
16	34.4	5.1	579	2	US-08-469-277-3
17	34.4	5.1	579	3	US-08-468-946-3
18	34.4	5.1	579	4	US-08-468-942-3
C 19	34.4	5.1	736	3	US-08-540-406-1
C 20	34.4	5.1	736	5	US-08-656-055-1
C 21	34.4	5.1	736	6	PCT-US95-13233-1
C 22	34.4	5.1	1212	4	US-09-092-770-18
23	34.2	5.1	9421	3	US-08-370-319C-2
24	34.2	5.1	246240	3	US-08-724-394A-20
25	34.2	5.1	246240	3	US-08-724-394A-21
26	34.2	5.1	246240	3	US-08-724-394A-22
27	33.6	5.0	303	2	US-08-190-560-1

28	33.6	5.0	303	2	US-08-469-277-1	Sequence 1, Appli
29	33.6	5.0	303	3	US-08-468-946-1	Sequence 1, Appli
30	33.6	5.0	303	4	US-08-488-942-1	Sequence 1, Appli
31	33.6	5.0	1269	2	US-08-808-641-2	Sequence 2, Appli
32	33.6	5.0	1269	4	US-09-064-839-2	Sequence 2, Appli
33	33.6	5.0	1269	5	US-09-351-438-2	Sequence 2, Appli
C 34	33.6	5.0	15225	4	US-08-892-403A-2	Sequence 2, Appli
35	33	4.9	366	4	US-08-853-659A-11	Sequence 11, Appli
36	33	4.9	2260	1	US-07-794-393-3	Sequence 3, Appli
37	33	4.9	2260	1	US-08-001-711-3	Sequence 3, Appli
38	33	4.9	15512	4	US-08-853-659A-5	Sequence 5, Appli
C 39	33	4.9	15512	4	US-08-853-659A-8	Sequence 8, Appli
40	33	4.9	15512	4	US-08-853-659A-63	Sequence 63, Appli
C 41	33	4.9	15512	4	US-08-853-659A-66	Sequence 66, Appli
42	33	4.9	24701	4	US-08-853-659A-2	Sequence 2, Appli
C 43	33	4.9	24701	4	US-08-853-659A-3	Sequence 3, Appli
44	33	4.9	24701	4	US-08-853-659A-60	Sequence 60, Appli
C 45	33	4.9	24701	4	US-08-853-659A-61	Sequence 61, Appli

ALIGNMENTS

```

RESULT 1
US-08-469-667-6
; Sequence 6, Application US/08469667
; Patent No. 5733748
; GENERAL INFORMATION:
; APPLICANT: Yu, Guo-Liang
; TITLE OF INVENTION: Colon Specific Genes and Proteins
; NUMBER OF SEQUENCES: 24
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Carella, Byrne, Bain, Gilfillan, Cecchi,
; ADDRESS: Stewart & Olstein
; CITY: Roseland
; STATE: NJ
; COUNTRY: USA
; ZIP: 07068-1739
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,667
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Ferraro, Gregory D.
; REGISTRATION NUMBER: 36,134
; REFERENCE/DOCKET NUMBER: 325800-435
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201-994-1700
; TELEFAX: 201-994-1744
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 548 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..405
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 1..405
; US-08-469-667-6

```

Query Match 44.0%; Score 294.6; DB 2; Length 548;


```

Query Match          5.98; Score 39.4; DB 4; Length 1215;
Best Local Similarity 9.0%; Pred.No. 0.033; 355; Indels 0; Gaps
Matches 52; Conservative 169; Mismatches
0:

Qy 94 aaracnathmgnytnccnmngtgytngcnwscnccnaagarcathcargttnaarar 153
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 634 ANGCRAAATCYTGNAryTNGGNGCGRTADATYTCYTCNARYTTNSWNGCDATRAANARNS 575

Qy 154 tayaratgyggnynathaaarcentgyccngcnaaaytaytytgcnttyaaarathtgywsn 213
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 574 WNGTDTATNCCDATNARYTGNArcATRTTYTTRTDTATCTCTTYTGNGFNARCATPRAANC 515

Qy 214 ggngcngcnaaytngtngngccnaccnatgtytgytgaygmngnatgathatgwsnccn 273
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 514 KRTCEAPAAATCYTGNGCNARRARAANGYYTCNCKRTGNARNGTRTANACYTCRCANA 455

Qy 274 gtnaarayaaytngtngmgnggnytnaayathgcnyngtngnaayggcnacnngngcn 333
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 454 CYTCNARNARCARTCNARNARDATNSWNCKATYTGNGGYTCNARRTCNSWRTGNARNA 395

Qy 334 gtnyngngncaraargcnttygayatgtaywsngnggngaygtngatcaytyngttnaarit 393
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 394 CYTCRAARTGYTTRCFRGNACRTANCKNSWYTCYTTTNNARCATRTTNARCCANACYT 335

Qy 394 ytnaargarathcngngngngcnyngtngtngtngcngnwsntaygaycngngnccn 453
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 334 CYTINSWRCANCCCCANSWNARRTCNGGNARNGNSWNGGRTTDTAANARRTTYTTRA 275

Qy 454 aaratgaaygaygarwsnmgnaaryntntywsngayyytngngnwsntayqcnarcar 513
   | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 274 ANCKRTAETTTNGTRAACKNSWNAARTCNNSWNGTNCDDATYTCYTRTGTGNGNGTYTCDA 215

Qy 514 ytnagntymngaywsntggntnttyathgngcnaargayyytngngngnaarwsnccn 573
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 214 TDATATRCANGNSWDATNCCNCCNSNARNACNGNGGCCARCATTNCKDKATYTCRT 155

Qy 574 ttygarcarttytnaarayaawscnccnccnaaayaaarataaygarngtggcngarytn 633
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 154 AYTGRGTGYTTTNGTNNACYTCYTCNCKNCKYTTTNNACRTCYTGTGNGTNGTYTTCNKYT 95

Qy 634 ytnagaratggargngtgyatgccnccnaarccntty 669
   : : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 94 YTTNGCYTGDATDATYTGNGCYTCYTCGTGNGNSWYT 59

RESULT 5
US-08-608-241-3
; Sequence 3, Application US/08608241
; Patent No. 5747328
; GENERAL INFORMATION:
; APPLICANT: Donohue, Timothy J
; APPLICANT: Barber, Robert D
; APPLICANT: Witthuhn, Vernon
; TITLE OF INVENTION: MICROBIAL SYSTEM FOR FORMALDEHYDE
; TITLE OF INVENTION: SENSING AND REMEDIATION
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Pinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53703
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/608,241
; FILING DATE:

```

```
;
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 960296.93511
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1058 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Rhodobacter sphaeroides
; STRAIN: 2.4.1
; IMMEDIATE SOURCE:
; CLONE: pUI8017
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 221..895
;
US-08-608-241-3

Query Match          5.4%; Score 35.8; DB 2; Length 1058;
Best Local Similarity 27.3%; Pred. No. 0.36;
Matches 93; Conservative 47; Mismatches 201; Indels 0; Gaps 0;

QY 102 hmgnytncnmngtgytngcnwsnccnaaargathcargtnaaraartayaartg 161
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 151 CTTTGGCGGGTTATCGCGTCAGGCGGATGTGTAGCCTCGTGAAGACGATATGGGAC 210
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 162 yggnytnaahccntgycngcnaaytaytgygcnttyaarathtgywsngngcngc 221
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 211 CGCATGATGCAACCTTCGCCCGCTCCAGTCCGCGCTGATCGTCGAGCATCCGGT 270
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 222 naaytngtngnccnaccnatgtgytgygargaymgnatgathatgwsnccngtnaaraa 281
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 271 CTTCTGCGATCGCGTGTGATGAGGCTCAAGCGGTCGAGCCGCTGACCCACATCGAATC 330
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 282 yaaytngngmngngnytnaayathcnyntngtnaayggnacnagngngcngtntngg 341
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 331 CGCCGACCGGCTCGAGACCGCTTGCGCCGCTCGACCTGCGAGCCGGCTTCGATGTGGT 390
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 342 ncaaraargcnttygayatgtaywsngngngaygtnatgcayytngtnaarttytnaarga 401
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 391 GGTGCTGGATCTGAACCTGCGGACGCTGAGCGGCTGAGCGGCTGATCGATGTGGT 450
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 402 rathcngngngngcnyntngtntngtngcnwsntaygayg 442
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 451 GAGCCCGGGGGTGGCGGTGGTGTGTCTCTTCCCTCG 491
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
US-08-922-182-3
; Sequence 3, Application US/08922182
; Patent No. 5834300
; GENERAL INFORMATION:
; APPLICANT: Donohue, Timothy J
; APPLICANT: Barber, Robert D
; APPLICANT: Wittruh, Vernon
; TITLE OF INVENTION: MICROBIAL SYSTEM FOR FORMALDEHYDE
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 1 South Plinckney Street
; CITY: Madison
; STATE: WI
; COUNTRY: US
; ZIP: 53703
```

```
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/922,182
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/608,241
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Seay, Nicholas J
; REGISTRATION NUMBER: 27,386
; REFERENCE/DOCKET NUMBER: 960296.93511
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 608-251-5000
; TELEFAX: 608-251-9166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1058 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Rhodobacter sphaeroides
; STRAIN: 2.4.1
; IMMEDIATE SOURCE:
; CLONE: pUI8017
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 221..895
;
US-08-922-182-3

Query Match          5.4%; Score 35.8; DB 3; Length 1058;
Best Local Similarity 27.3%; Pred. No. 0.36;
Matches 93; Conservative 47; Mismatches 201; Indels 0; Gaps 0;

QY 102 hmgnytncnmngtgytngcnwsnccnaaargathcargtnaaraartayaartg 161
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 151 CTTTGGCGGGTTATCGCGTCAGGCGGATGTGTAGCCTCGTGAAGACGATATGGGAC 210
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 162 yggnytnaahccntgycngcnaaytaytgygcnttyaarathtgywsngngcngc 221
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 211 CGCATGATGCAACCTTCGCCCGCTCCAGTCCGCGCTGATCGTCGAGCATCCGGT 270
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 222 naaytngtngnccnaccnatgtgytgygargaymgnatgathatgwsnccngtnaaraa 281
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 271 CTTCTGCGATCGCGTGTGATGAGGCTCAAGCGGTCGAGCCGCTGACCCACATCGAATC 330
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 282 yaaytngngmngngnytnaayathcnyntngtnaayggnacnagngngcngtntngg 341
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 331 CGCCGACCGGCTCGAGACCGCTTGCGCCGCTCGACCTGCGAGCCGGCTTCGATGTGGT 390
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 342 ncaaraargcnttygayatgtaywsngngngaygtnatgcayytngtnaarttytnaarga 401
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 391 GGTGCTGGATCTGAACCTGCGGACGCTGAGCGGCTGAGCGGCTGATCGATGTGGT 450
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 402 rathcngngngngcnyntngtntngtngcnwsntaygayg 442
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 451 GAGCCCGGGGGTGGCGGTGGTGTGTCTCTTCCCTCG 491
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 7
US-08-919-953-3
; Sequence 3, Application US/08919953
; Patent No. 5837481
; GENERAL INFORMATION:
; APPLICANT: Donohue, Timothy J
; APPLICANT: Barber, Robert D
```



```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/190,560
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Digiglio, Frank S.
; REGISTRATION NUMBER: 31,346
; REFERENCE/DOCKET NUMBER: 7879ZY
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (516) 742-4343
; TELEFAX: (516) 742-4366
; TELEX: 230 901 SANS UR
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 579 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-190-560-3

```

```

Query Match      5.1%; Score 34.4; DB 2; Length 579;
Best Local Similarity 27.8%; Pred. No. 0.57;
Matches 72; Conservative 40; Mismatches 147; Indels 0; Gaps 0;

Qy 350 cnttygayatgtaywsgnggaygtatgcaytyngtngaarttyytnaaragarathccng 409
Db 179 CTTTCCACAAGTACTCGGGCAAGAGGGGTGACAAAGTTCAAGCTCAACAGTCAAGTCAAGACTAA 238
Qy 410 gngngcnytngtngtngcnwsntaygayccnggnacnaaratgaaygaygarw 469
Db 239 AGGAGCTCTGACCCGGGAGCTGCCAGCTTCTTGGGGAAAAGGACAGATGAAGCTGCTT 298
Qy 470 snmgnaarynttywsgngaytyngnwnsnwtaygcnarcarrytngnttymgngayw 529
Db 299 TCCAGAAGCTGATGAGCAACTTGGACACACAGGAGGACACAGGAGGTGGACTTCCAAGAGT 358
Qy 530 sntgggtnttythggngcnaaragaytnmngngnnaarwsnccnttygarcarttyyna 589
Db 359 ACTGTGTCTTCTGTCTGTCATCGCCATGATGTGTAACGAATCTTTGAAGGCTTCCAG 418
Qy 590 araaywscncngayacnaa 608
Db 419 ATAAGCAGCCCGAGGAAGAA 437

```

Search completed: May 31, 2000, 19:14:25
Job time: 6263 sec

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	485.8	72.6	1299	1	V63516	Consensus sequence
2	447	66.8	955	1	V63515	Full length sequen
3	294.6	44.0	548	1	T45883	Human colon specif
4	294.6	44.0	548	1	V16671	Polynucleotide seq
5	193.6	28.9	263	1	V63509	Clone 1281865 EST
6	181.4	27.1	251	1	V63511	Clone 1735382 EST
7	161.8	24.2	228	1	V63510	Clone 1807758 EST
8	59.2	8.8	287	1	V63508	Clone 3210629 EST
9	40.4	6.0	187	1	V63512	Clone 2128334 EST
10	39	5.8	3830	1	N70938	DNA fragment contg
11	37.8	5.7	7041	1	V74910	Staphylococcus aur
12	37	5.5	1341	1	V69614	Human secreted pro
13	36.8	5.5	1633	1	T08144	Xylanase xynB gene
14	36.6	5.5	987	1	V31480	Human interleukin-
15	36.6	5.5	3934	1	T30309	Human SOX-9 cDNA.
16	36.4	5.4	8536	1	V74589	Staphylococcus aur
17	36.4	5.4	110000	1	V21209_09	Continuation (10 o
18	35.8	5.4	1058	1	V62126	R. sphaeroides Adh
19	35.8	5.4	1058	1	V64113	Rhodobacter sphaer
20	35.6	5.3	1836	1	N20057	Interferon-beta ge
21	35.6	5.3	4544	1	V74828	Staphylococcus aur
22	35.4	5.3	1836	1	T29734	Brevibacterium fla
23	35.4	5.3	2262	1	V06554	Alcaligenes (Deley
24	35.2	5.3	3101	1	V32918	Solanum tuberosum
25	35	5.2	420	1	V77877	Staphylococcus aur
26	35	5.2	1442	1	T83899	DNA encoding a Sta
27	35	5.2	1442	1	V53367	DNA encoding a Sta
28	35	5.2	2855	1	V00011	Human sex comb on
29	35	5.2	2855	1	T97297	Human sex comb on
30	35	5.2	3255	1	T97299	Human sex comb on
31	35	5.2	3255	1	V00013	Human sex comb on
32	35	5.2	3327	1	T97298	Human sex comb on
33	35	5.2	3327	1	V00012	Human sex comb on
34	35	5.2	14078	1	V74502	Staphylococcus aur

Qy	181	ccngcnaaytaytytgnttyaarathgywsngngcngcnaaygtngtngncncaacn	240
Db	466	CCAGCCAACTACTTTGGCTTTAAAACTGCAGTGGGCGCCCAACGTCGTGGGCCCTACT	525
Qy	241	atgtgytytgargaymgnatgathatgwsncngttnaarayaaygtngnmgngnytn	300
Db	526	ATGTGCTTTGAAGACCCGATGATCATGAGTCCGTGAAAAACAATGTGGCAGAGGCCTA	585
Qy	301	aayathocnnytnaaygnacnngcngngtngtngncncaraargonttygyatg	360
Db	586	AACATCCGCTTGGTGAATGAACACACGGGAGCTGTGCTGGACAGAAAGCATTTGCATG	645
Qy	361	taywsngngaygtnatgcyvytngtnaarttytnaaragarathcngngngngcnytn	420
Db	646	TACTCTGGAGATGTTATGCACCTAGTGAATTCCTTAAGAAATTCGGGGGGTGCACTG	705
Qy	421	gtntyngtngcnwsntaygaygcngngnacnaaratgaaygaygarwsnmgnaarytn	480
Db	706	GTGCTGTGGCTCTACGACGATCCAGGGACCAAAATGAACGATGAAGACAGGAAACTC	765
Qy	481	ttywsngaytyngnwnswntaygcnaacarytngntngtymngaywsntgggtntty	540
Db	766	TTCTCTGACTGGGGAGTTCCTACGCCAAAAACAATGGGCTTCGGGACAGCTGGGTCTTC	825
Qy	541	athgngcnaargaytynmngnaarwsncnttygarcttytnaarayawsnccn	600
Db	826	ATAGAGCCCAAGACCTCAGGGGTAAAGGCCCTTTGAGCAGTCTTAAGAACAACGCCCA	885
Qy	601	gayacnaaartaytargnttggcngarytytnyngaratgargngtgyatgcncncn	660
Db	886	GACACAAACAAATACGAGGGATGSCCAGACTCTCTGGAGATGAGGGCTGCATGCCCCCG	945
Qy	661	aarcnnty 669	
Db	946	AAGCCATTT 954	

RESULT	2	
V63515	standard; cDNA; 955 BP.	
ID	V63515;	
AC	28-JAN-1999 (first entry)	
DT	Full length sequence of clone 1281865 corresponding to CS141 sequence.	
DE	EST sequence; CS141: gastrointestinal tract; cancer; Barrett's esophagus;	
KW	gastric ulcer; gastritis; leiomyoma; polyps; Crohn's disease;	
KW	ulcerative colitis; pancreatitis; ss.	
OS	Synthetic.	
OS	Homo sapiens.	
PN	WO9844133-A1.	
PD	08-OCT-1998.	
PF	31-MAR-1998; U06337.	
PR	31-MAR-1997; US-828489.	
PA	(ABBO) ABBOTT LAB.	
PI	Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,	
PI	Gordon J, Granados EN, Hayden M, Hodges SC, Klass MR,	
PI	Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;	
DR	WPI; 98-568280/48.	
PT	New gastrointestinal tract specific polynucleotides, CS141 - used to	
PT	develop products for the diagnosis and treatment of e.g. cancers,	
PT	gastric ulcer, gastritis, Crohn's disease, ulcerative colitis or	
PT	pancreatitis	
PS	Claim 1; Page 92; 116pp; English.	
CC	V63504-15 represent a set of contiguous and partially overlapping EST	
CC	sequences designated CS141. The sequences are isolated from a cDNA	
CC	library made from gastrointestinal tract tumour and normal tissues.	
CC	The CS141 gene is useful as a marker for gastrointestinal tract	
CC	disorders. The methods and products can be used in detecting,	
CC	diagnosing, staging, monitoring, prognosticating, preventing, or	
CC	determining the predisposition to diseases and conditions of the	
CC	gastrointestinal tract, such as gastrointestinal tract cancers,	
CC	Barrett's esophagus, gastric ulcer, gastritis, leiomyoma, polyps,	
CC	Crohn's disease, ulcerative colitis, pancreatitis.	

Seq	Sequence	955 BP;	241 A;	240 C;	273 G;	197 T;
Query Match						
Best Local Similarity 66.8%; Score 447; DB 1; Length 955;						
Matches 366; Conservative 141; Mismatches 103; Indels 0; Gaps						
Qy	60	gatgttathmgnsntayatsnttysnatgaaracnathmngyntncnmngntggyt	119			
Db	1	GATGTTTATTCGAAGCTACATGAGCTTCAGCATGAAACCATCCGCTGCCACGCTGGCT	60			
Qy	120	ngcnswcnacnaaragarathcaarthaarataartayaaartgyggnynathaaarcentg	179			
Db	61	GGCCTCGCCACCAAGGAGATCCAGGTATAAAATGACAGTGTGGCTCATCAAGCCCTG	120			
Qy	180	ycngcnaaytytycncntyaarathctgywsngngcngcnaayvtngtngngcncac	239			
Db	121	CCGACCCAACTACTTTTCGGTTTAAATCTCAGTGGGGCCGCCAAGCTGTGGGCCCTAC	180			
Qy	240	natgtvttytgargaymgntgathatgwnccngnnaaraayaavtngngmngngnytt	299			
Db	181	TATGTGCTTTGAAGACCGCATGATCATGAGTCTCTGTCAAAACAATGTGGCAGAGCCT	240			
Qy	300	naayathgcnytngtnaayggnacnccngngcngtnytnngncaraaragcncnttygayt	359			
Db	241	AAACATCGCCTGGTGAATGGAACCCACGGGAGCTGTCTGGGACAGAAGGCATTTGCAT	300			
Qy	360	qtaywsngngaytgatnatcayvtngtnaarttytnaargarathccngngngcngcnyt	419			
Db	301	GTACTCTGGAGATGTTATGACACCTAGTGTAAATTCCTTAAAGAAATTCGGGGGGTGCAC	360			
Qy	420	ngtnytngtngcwnstaygaycaycngcngnacnaaratgaaygarwsmngnaaryt	479			
Db	361	GGTCTGTGGTCCCTACAGACGATCCAGGGACCAAAATGACATGAAGAAGCAGGAAACT	420			
Qy	480	nttysngaytytngnwsnwsntaygcnacrcarytngnttymngngaywsntgggtntt	539			
Db	421	CTTCTCTGACTTGGGAGTTCCTACGCAAAACAACATGGGCTTCCGGGACAGCTGGGCTTT	480			
Qy	540	yathgngcnaargaytymngngnnaarwsnccnttygarcarttytnaaraaywsncc	599			
Db	481	CATAGGACCCAAAGACCTCAGGGGTAAAAGCCCTTTGAGCAGTTTCTTAAAGAACAACGCC	540			
Qy	600	ngayacnaayaartaygargntggcngarytnytnngaratgagtgntgyatgcnc	659			
Db	541	AGACACAACAATACAGCGGATGGCCAGAGCTGCTGGAGATGGAGGGCTGCATGCCCCC	600			
Qy	660	naarcentg 669				
Db	601	GAAGCCATT 610				
RESULT 3						
T45883						
ID	T45883 standard; cDNA; 548 BP.					
AC	T45883;					
DT	13-MAR-1997 (first entry)					
DE	Human colon specific gene CSG4 cDNA partial clone.					
KW	Colon specific gene; CSG4; colon cancer; metastasis; diagnosis;					
KW	gene therapy; ss.					
OS	Homo sapiens.					
FH	Key					
FT	cds					
FT	1..408					
PN	/*tag- a					
PD	W09639419-A1.					
PN	12-DEC-1996.					
PF	06-JUN-1995; U07289.					
PR	06-JUN-1995; W07289.					
PA	(HUMA-) HUMAN GENOME SCI INC.					
PI	Rosen CA, Yu G;					
DR	WPI; 97-043054/04.					
DR	P-PSDB; W06547.					
PT	Human colon specific genes and their expression products - detection					

PT of which, in non-colon tissue samples, can be used as indication of
PS colon cancer metastasis
PS Claim 1; Fig 4; 60pp; English.
CC 13 cDNA clones (T45880-92), most of them partial clones, correspond
CC to human colon specific genes, designated CSG1, CSG2, etc., that
CC are primarily expressed in tissues derived from the colon. CSG7
CC and CSG10 show reduced expression in colon cancer cells as compared
CC to that in normal cells; the remaining genes are overexpressed in
CC colon cancer. The partial cDNA sequences can be used to isolate
CC full-length clones and genomic clones including the complete gene.
CC CSG nucleic acids can be used to produce CSG polypeptides (see also
CC W06545-53) in transformed host cells, as probes to detect disorders
CC of the colon, partic. colon cancer and colon cancer metastasis, and
CC in gene therapy.
SQ Sequence 548 BP; 137 A; 129 C; 172 G; 107 T;

Query Match 44.0%; Score 294.6; DB 1; Length 548;
Best Local Similarity 59.3%; Pred. No. 2.8e-75;
Matches 240; Conservative 93; Mismatches 72; Indels 0; Gaps 0;
QY 265 atgwsncngntnaaraaayagtngngmngngnytnaayathgcnynytngnaaygnacn 324
DB 1 ATGAGTCTCTGTGAACAACATGTGGCAGAGGCGCTAAACATCGCCCTGGTGAATGGAACC 60
QY 325 acngngngcngntnngncaraargcnttygayatgtaywsngngngaygtngatgcaytyn 384
DB 61 ACGGGAGCTGTGCTGGGACAGAGGCATTGTGACATGTACTCTGGAGATGTATGACACCTA 120
QY 385 gtnaarttytnaargathccngngngcngntnngntnngntnngcwnstnngaygay 444
DB 121 GTGAAATTCCTTTAAAGAAATTCGGGGGGTGCACCTGGTGTGGCTCTCTACGACGAT 180
QY 445 ccngnnaacnaaratgaaygagaygarwsnmgnaaryntnttyathgngcnaargaytngmngn 504
DB 181 CCAGGGACCAAAATGAACATGAAGCAGGAACTCTCTGACTGTGGCAGTCTCTAC 240
QY 505 gnaarcarytngnttymngaywsntgggtnttyathgngcnaargaytngmngn 564
DB 241 GCAAAACAACACTGGGCTTCGGGACAGCTGGTCTTCATAGAGCCAAAGACCTCAGGGGT 300
QY 565 aarwsncnttygarcarttytnaaraaywsncngaycnaayartaaygargntgg 624
DB 301 AAAAGCCCCCTTTGAGCAGTCTTTAAAGAACAGCCGACAGACAAACAAATACGAGGATGG 360
QY 625 ccngarytngatgargntgagntgyatgccnccnaarcentty 669
DB 361 CCAGAGCTGTGGAGTGGAGGCTGCATGCCCCCGAAGCCATT 405

RESULT 4
V16671
ID V16671 standard; cDNA; 548 BP.
AC V16671;
DE 22-JUN-1998 (first entry)
DE Polynucleotide sequence of a colon-specific gene.
KW Colon-specific gene; probe; detection; expression; human;
KW diagnostic assay; colon cancer; antibody; screening; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT CDS 1..406
FT /*tag= a
FT US5733748-A.
PN 31-MAR-1998.
PD 06-JUN-1995; 459667.
PR 06-JUN-1995; US-459667.
PA (HUMA-) HUMAN GENOME SCI INC.
PI Rosen C, Yu G;
DR WPI; 98-229823/20.
DR P-PSDB; W46878.
PT Colon-specific nucleic acids - useful as probes for detecting colon
PS cancer micrometastases
PS Claim 15; Fig 4; 51pp; English.

CC V16668-81 represent polynucleotide sequences of partial or full-length
CC cDNA clones of colon-specific genes. The polynucleotides can be used
CC as probes to detect expression of the corresponding human genes, e.g. in
CC diagnostic assays for detecting micrometastases of colon cancer.
CC Recombinant cells containing the polynucleotides can be used to
CC produce the polypeptides, in order that antibodies can be raised and
CC used in further screening or diagnostics.
SQ Sequence 548 BP; 137 A; 128 C; 173 G; 107 T;
Query Match 44.0%; Score 294.6; DB 1; Length 548;
Best Local Similarity 59.3%; Pred. No. 2.8e-75;
Matches 240; Conservative 93; Mismatches 72; Indels 0; Gaps 0;
QY 265 atgwsncngntnaaraaayagtngngmngngnytnaayathgcnynytngnaaygnacn 324
DB 1 ATGAGTCTCTGTGAACAACATGTGGCAGAGGCGCTAAACATCGCCCTGGTGAATGGAACC 60
QY 325 acngngngcngntnngncaraargcnttygayatgtaywsngngngaygtngatgcaytyn 384
DB 61 ACGGGAGCTGTGCTGGGACAGAGGCATTGTGACATGTACTCTGGAGATGTATGACACCTA 120
QY 385 gtnaarttytnaargathccngngngcngntnngntnngntnngcwnstnngaygay 444
DB 121 GTGAAATTCCTTTAAAGAAATTCGGGGGGTGCACCTGGTGTGGCTCTCTACGACGAT 180
QY 445 ccngnnaacnaaratgaaygagaygarwsnmgnaaryntnttyathgngcnaargaytngmngn 504
DB 181 CCAGGGACCAAAATGAACATGAAGCAGGAACTCTCTGACTGTGGGAGTCTCTAC 240
QY 505 gnaarcarytngnttymngaywsntgggtnttyathgngcnaargaytngmngn 564
DB 241 GCAAAACAACACTGGGCTTCGGGACAGCTGGTCTTCATAGAGCCAAAGACCTCAGGGGT 300
QY 565 aarwsncnttygarcarttytnaaraaywsncngaycnaayartaaygargntgg 624
DB 301 AAAAGCCCCCTTTGAGCAGTCTTTAAAGAACAGCCGACAGACAAACAAATACGAGGATGG 360
QY 625 ccngarytngatgargntgagntgyatgccnccnaarcentty 669
DB 361 CCAGAGCTGTGGAGTGGAGGCTGCATGCCCCCGAAGCCATT 405
RESULT 5
V63509
ID V63509 standard; cDNA; 263 BP.
AC V63509;
DE 28-JAN-1999 (first entry)
DE Clone 1281865 EST corresponding to CS141 sequence.
KW EST sequence; CS141; gastrointestinal tract; cancer; Barret's esophagus;
KW gastric ulcer; gastritis; leiomyoma; polyps; Crohn's disease;
KW ulcerative colitis; pancreatitis; ss.
OS Synthetic.
OS Homo sapiens.
PN W09844133-A1.
PD 08-OCT-1998.
PF 31-MAR-1998; U06337.
PR 31-MAR-1997; US-828489.
PA (ABDO) ABBOTT LAB.
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
PI Gordon J, Granados EN, Hayden M, Hodges SC, Klass MR,
PI Kratochvil JD, Roberts-Rapp L, Russell JC, Stroupe SD;
DR WPI; 98-568280/48.
PT New gastrointestinal tract specific polynucleotides, CS141 - used to
PT develop products for the diagnosis and treatment of e.g. cancers,
PT gastric ulcer, gastritis, Crohn's disease, ulcerative colitis or
PT pancreatitis
PS Claim 1; Pages 90-91; 116pp; English.
CC V63504-15 represent a set of contiguous and partially overlapping EST
CC sequences designated CS141. The sequences are isolated from a cDNA
CC library made from gastrointestinal tract tumour and normal tissues.
CC The CS141 gene is useful as a marker for gastrointestinal tract
CC disorders. The methods and products can be used in detecting,


```
PF 31-MAR-1998; U06337.
```

```
PR 31-MAR-1997; US-828489.
```

```
PA (ABBO ) ABBOTT LAB.
```

```
PI Billing-Medel PA, Cohen M, Colpitts TL, Friedman PN,
```

```
PJ Gordon J, granados EN, Hayden M, Hodges JC, Klass MR,
```

```
PK Kratochvil JD, Roberts-Rapp L, Russell SC, Stroupe SD;
```

```
DR WPI; 98-568280/48.
```

```
PT New gastrointestinal tract specific polynucleotides, CS141 - used to
```

```
PP develop products for the diagnosis and treatment of e.g. cancers,
```

```
PT gastric ulcer, gastritis, Crohn's disease, ulcerative colitis or
```

```
PT pancreatitis
```

```
PS Claim I.; Pages 91-92; l16pp; English.
```

```
CS V03504-15 represent a set of contiguous and partially overlapping ES-
```

```
CC sequences designated CS141. The sequences are isolated from a cDNA
```

```
CC library made from gastrointestinal tract tumour and normal tissues.
```

```
CC The CS141 gene is useful as a marker for gastrointestinal tract
```

```
CC disorders. The methods and products can be used in detecting,
```

```
CC diagnosing, staging, monitoring, prognosticating, preventing, or
```

```
CC determining the predisposition to diseases and conditions of the
```

```
CC gastrointestinal tract, such as gastrointestinal tract cancers,
```

```
CC Barrett's esophagus, gastric ulcer, gastritis, leiomyoma, polyyps,
```

```
CC Crohn's disease, ulcerative colitis, pantiectinitis.
```

```
SQ Sequence 187 Bp;   35 A;    42 C;    74 G;    33 T;
```

```
Query Match           6.0%; Score 40.4; DB 1; Length 187;
```

```
Best Local Similarity 67.3%; Pred.No.0.018;
```

```
Matches 35; Conservative 9; Mismatches 8; Indels 0; Gaps
```

```
QY      618 rgngtgccgagarynnyntgaratggarggtgatgcacncnaarcentty 669  
       :|||:|||::|||::|||::::|||||::|||::::|||::|||::|||:
```

```
Dd      1 GGATGCCAGACGTCGTGGAGATGAGGCGTGCATGCCCCCAGAACCATT 52
```

```
RESULT 10
```

```
N70938
```

```
ID N70938 standard; DNA; 3830 BP.
```

```
IC N70938;
```

```
DT 10-APR-1991 (first entry)
```

```
DE DNA fragment conty. streptomycin resistance gene and its regulatory
```

```
DE sequence from Streptomyces griseus 4-1 strain (FERM BP-1198).
```

```
KW Antibiotic-resistance; selection marker; ss.
```

```
KS Streptomyces griseus.
```

```
FH Key Location/Qualifiers
```

```
FT cds          1043..1966
```

```
FT /*tag= a
```

```
PD WO8703302-A.
```

```
PN 04-JUN-1987.
```

```
PR 20-NOV-1986; J00596.
```

```
PR 20-NOV-1985; JP-258622.
```

```
PA (ZAID.) ZH BISETUTSU KAKAKU KEN.
```

```
PA (UMEZ.) UMEZAWA H.
```

```
PI Umezawa H, Ikami Y;
```

```
PJ WPI; 87-163322/23.
```

```
DP DNA contg. streptomycin resistance gene and regulatory sequence -
```

```
PT useful as resistance selection marker, isolated from plasmid
```

```
PT pST141 Streptomyces lividans 4-1
```

```
PS Disclosure; Fig 3; 4lp; Japanese.
```

```
CC The DNA fragment can be inserted into an actinomycetes plasmid
```

```
CC vector (e.g. pIJ720) to give a hybrid plasmid which may be used as a
```

```
CC streptomycin resistance selection marker.
```

```
SQ Sequence 3830 Bp; 576 A; 1367 C; 1348 G; 538 T;
```

```
Query Match           5.8%; Score 39; DB 1; Length 3830;
```

```
Best Local Similarity 27.0%; Pred.No.0.46;
```

```
Matches 89; Conservative 49; Mismatches 192; Indels 0; Gaps
```

```
QY     195 ycntttaaathctgywsngcngcnuaavgtngtngccnacnatgttttytgarga 250  
       :|||:|||::|||::|||::|||::|||::|||::|||::|||::|||:
```

```
Dd     17 CGCCCAAAATGCGCGGAGCCCGCATGCGCTGCGGTATTTCGAAGCGGAGAGA 76
```


CC useful for preventing, treating or ameliorating medical conditions e.g.
CC by protein of gene therapy. Also pathological conditions can be diagnosed
CC by determining the amount of the new polypeptides in a sample or by
CC determining the presence of mutations in the polynucleotides. Specific
CC uses are based on which tissues they are most highly expressed in, and
CC include developing products for the diagnosis or treatment of central
CC nervous system (CNS) and immune system diseases, reproductive disorders,
CC cancers, congenital malformations, degenerative diseases, trauma,
CC inflammatory disease, neoplasia, metabolic disorders, diseases in testes,
CC placenta, liver, brain and activated T cells, spleen diseases, lung
CC diseases, heart diseases, rhabdomyosarcoma and disorders of the endocrine
CC system or other endocrinopathies, e.g. endocrine polyglandular syndrome,
CC endocrinoma, and endocrine ophthalmopathy, osteoclastoma and other bone
CC remodelling disorders, bacterial infections and sepsis. The polypeptides
CC are also useful for identifying their binding partners.
SQ Sequence 1341 BP; 376 A; 278 C; 313 G; 373 T;

Query Match 5.5%; Score 37; DB 1; Length 1341;
Best Local Similarity 28.3%; Pred. No. 0.77;
Matches 91; Conservative 60; Mismatches 167; Indels 4; Gaps 2;

QY 282 yaaygtngngmgngnytnaayathgcnyngtngtnaaygngnacngngcngntngng 341
DB 553 CAACGTAGGAAGAGAGACAGAAACATCTCTCTCCACCTTATTACCTGCCCTGTGTT--- 496
QY 342 ncaaraagcnttygatytaywsngngngaygtnatgcaytngtnaarttytnaarga 401
DB 496 TTAAGAGGTATATGGTATTAGAAAGATGGCTGTGCTGTTTAAATTTTAAATTT 437
QY 402 rathccngngngcnytngtngtngncnwsntaygayccngncaaraatgaa 461
DB 436 ATGAACTAGCGCTGCCACTCTCTCTGCTTACTTGTAGGTGATCTAGCTGCC 377
QY 462 ygaygarwsnmgnaaryntnt-ywsngaytngngnwsntaygnaacarytngngat 520
DB 376 TGTGGTATCATATTAGGATCCCATCAATGATTGGATAAGCTATTCCCAACTCTTCATTA 317
QY 521 tynngaywsntgggtnttyathgngnaargaytngngngnaarwsnccnttygarc 580
DB 316 ATCAATTCGTTTGTGTTGATCTATATCTGAGCGGCTCTTGGAGAGCGGCACACGAG 257
QY 581 arttytnaaraaywsnccnga 602
DB 256 AACTCCAGCAGCGCGGATCGA 235

RESULT 13
T08144
ID T08144 standard; DNA; 1633 BP.
AC T08144;
DT 12-MAY-1996 (first entry)
DE Xylanase xynB gene (partial sequence).
KW Xylanase B; xynB gene; thermostable enzyme; paper; pulp; lignin;
KW delignification; xylan; bleaching; ds.
OS Extremophile isolate TG456 (CBS 213.94).
PN W09534662-A1.
PF 21-DEC-1995.
PD 14-JUN-1995; E02299.
PR 14-JUN-1994; EP-201699.
PA (KONN) GIST-BROCADES BV.
PI Bergquist PL, Daniel RM, Farrel RL, Forster S, Gronberg V;
PI Herweller MA, Iverson S, Jones BE, Moody D, Morgan HW;
PI Quax WJ, Williams DP;
DR WPI; 96-049690/05.
DR P-PSDB; R87014.
PT Novel xylanase(s) having activity at 80 deg C. or higher - obtained
PT from anaerobic thermophilic bacteria and used in paper and pulp
PT production processes
PS Example 8; Page 47-49; 77pp; English.
CC A xynB gene partial sequence (T08144) codes for a thermostable F-type
CC xylanase (R87014) useful in the pulp and paper industries. The
CC xynB gene, which codes for a multi-domain enzyme that includes a

CC single xylanase B domain, was obtd. by genomic walking PCR of DNA from
CC extreme thermophilic strain TC456 (CBS 213.94), isolated from a New
CC Zealand hot spring. The gene can be inserted into a vector and used
CC for the prodn. of recombinant xylanase B in microbial host cells, esp.
CC Escherichia coli.
SQ Sequence 1633 BP; 564 A; 291 C; 347 G; 431 T;

Query Match 5.5%; Score 36.8; DB 1; Length 1633;
Best Local Similarity 27.9%; Pred. No. 1;
Matches 57; Conservative 41; Mismatches 106; Indels 0; Gaps 0;

QY 449 gnacnaaratgaaygaygarwsnmgnaarynttywsngaytngngnwsntaygna 508
DB 409 GCACAGACCAATATATTTGGATTAAGAGGTCTATACACTGGTTGGCATCAACAACCTCCA 468
QY 509 arcarytngngntymngngaywsntgggtnttyathgngncnaargaytngngngnaarw 568
DB 469 GATTGGTTTTTCCAGCATCTCTGACGGTTCGCCACTTGATCCAAACAATTCCTGAAGACAAG 528
QY 569 snccnttygarcarttytnaaraaywsnccngayacnaayaartaaygarngntggcng 628
DB 529 CAGCTTTTTCAGAAAATAGGTTAAACACACATTCAGACACTTGTGTGGAAGATATGCGACAG 588
QY 629 acrytngngaratgargngtgya 652
DB 589 AAAGTTTATGATGGGATGTTGTA 612

RESULT 14
V31480
ID V31480 standard; DNA; 987 BP.
AC V31480;
DT 08-SEP-1998 (first entry)
DE Human interleukin-12 (IL-12) p40 subunit encoding DNA.
KW Interleukin-12; IL-12; p40 subunit; p35 subunit; human; asthma;
KW cancer; therapeutic gene delivery; vaccine; allergy; ss.
OS Homo sapiens.
FH Key Location/Qualifiers
FT Key 1. 987
FT CDS
FT /product= "human IL-12 p40 subunit"
FT /note= "the protein is encoded by degenerate codons
FT which can code for each of the amino acid"

W09817814-A2.
PD 30-APR-1998.
PF 10-OCT-1997; U18832.
PR 18-OCT-1996; US-028687.
PA (GENE-) GENEMEDICINE INC.
PI Deshpande D, Freimark B, Nordstrom J;
DR WPI; 98-261507/23.
DR P-PSDB; W57482.
PT - useful for, e.g. therapeutic gene delivery, especially for
PT treating asthma and cancer
PS Disclosure; Fig 4A-D; 104pp; English.
CC This DNA encodes a human interleukin-12 (IL-12) p40 subunit. This is used
CC in the construction of plasmid for expressing recombinant eukaryotic
CC genes. The plasmid comprises a first transcription unit comprising
CC control sequences linked, in order, to a 5'-untranslated region (UTR),
CC synthetic intron, a coding sequence and synthetic 3'-untranslated region
CC /polyA sequence (UTR/A), and a second transcription unit of similar
CC construction. The first coding sequence and second coding sequence
CC comprise a sequence coding for p40 subunit of human IL-12 and a sequence
CC coding for p35 subunit of human IL-12. The plasmids are used for
CC therapeutic gene delivery, especially for treating asthma and cancer.
CC They are also useful as adjuvants for vaccines. IL-12 is known to
CC stimulate natural killer cells, to promote maturation of cytotoxic T
CC cells and to induce cytokine production by these cells. They stimulate
CC cell-mediated immune responses in cases of microbial or viral infection,
CC some cancers, allergies and asthma. Co-ordinated expression of two
CC subunits from a single plasmid is now possible, avoiding the difficulties
CC associated with transfection with two separate plasmids.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 31, 2000, 18:37:16 ; Search time 1292.59 Seconds
(without alignments)
-503.483 Million cell updates/sec

Title: US-09-186-342-8
Perfect score: 669
Sequence: 1 atmgngtwnsgngngtntyt.....gyatgccnccnaarcntty 669

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 882769 seqs, -486395729 residues
Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

GenEmbl:*					Result					Query					Description				
1:	2:	3:	4:	5:	No.	Score	Match	Length	DB	ID	%								
1: gb_ba1.*	2: gb_ba2.*	3: gb_ov.*	4: gb_ov.*	5: gb_pat.*	1	294.6	44.0	548	5	I95745									
6: gb_ph.*	7: gb_pl1.*	8: gb_pl2.*	9: gb_pl1.*	10: gb_pr2.*	2	260.6	39.0	1606	9	HS219MRNA									
11: gb_pr3.*	12: gb_ro.*	13: gb_sts.*	14: gb_sy.*	15: gb_un.*	3	248.2	37.1	2475	9	D87120									
16: gb_vi.*	17: em_fun.*	18: em_hum1.*	19: em_hum2.*	20: em_in.*	c 4	189.8	28.4	465	12	MMU72677									
21: em_om.*	22: em_ov.*	23: em_ph.*	24: em_ph.*	25: em_pl.*	c 5	90.8	13.6	178137	55	AC023346									
26: em_ro.*	27: em_ro.*	28: em_sts.*	29: em_sy.*	30: em_un.*	c 6	76.2	11.4	52173	9	HSG6PDCEN									
31: em_vi.*	32: gb_htg1.*	33: gb_htg2.*	34: gb_in1.*	35: gb_in2.*	c 7	76.2	11.4	219447	40	HUMFLNG6PD									
36: em_ba1.*	37: em_ba2.*	38: em_hum3.*	39: em_hum4.*	40: gb_pr4.*	c 8	61.4	9.2	132319	40	AC006364									
41: gb_htg3.*	42: gb_htg4.*	43: gb_htg5.*	44: gb_htg6.*		c 9	52	7.8	37777	34	CELM70									
					c 10	52	7.8	151727	41	AC006894									
					c 11	48.6	7.3	7218	5	I66494									
					c 12	48.2	7.2	1302	51	AF090946									
					c 13	48	7.2	6031	1	CBGIDPAB									
					c 14	44	6.6	22877	32	AL133377									
					c 15	43.2	6.5	249287	41	AC006745									
					c 16	42.6	6.4	69214	43	AC018572									
					c 17	41.8	6.2	61771	54	AC023101									
					c 18	41.6	6.2	67994	40	AC006981									
					c 19	41.6	6.2	74371	11	AC005369									
					c 20	41.6	6.2	175986	42	AC010126									
					c 21	41.2	6.2	5527	34	TETTRAG									
					c 22	41.2	6.2	84630	45	AC008933									
					c 23	41.2	6.2	136862	40	AC005007									
					c 24	41.2	6.2	183516	44	AC018368									
					c 25	41	6.1	1409	8	AF016100									
					c 26	41	6.1	12203	1	SGSTR									
					c 27	41	6.1	19292	34	CEY76A2A									
					c 28	41	6.1	269082	32	CEY76A2									
					c 29	40.6	6.1	4161	7	ATHRD21A									
					c 30	40.6	6.1	34598	35	CELT09D3									
					c 31	40.6	6.1	41229	34	CEC15H11									
					c 32	40.6	6.1	64168	52	AC008607									
					c 33	40.6	6.1	79655	44	AC008724									
					c 34	40.6	6.1	133599	44	AC011339									
					c 35	40.6	6.1	181923	52	AC011741									
					c 36	40.6	6.1	226574	32	CEY49A3									
					c 37	40.2	6.0	2113	34	AB017107									
					c 38	39.8	5.9	153507	43	AC015823									
					c 39	39.8	5.9	165766	58	AC018361									
					c 40	39.8	5.9	169889	44	AC016356									
					c 41	39.8	5.9	171967	44	AC011678									
					c 42	39.8	5.9	191660	33	AL138718									
					c 43	39.6	5.9	99670	40	AC007200									
					c 44	39.6	5.9	99670	41	AC008677									
					c 45	39.6	5.9	203662	42	AC011235									

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

ALIGNMENTS


```
RESULT 3
D87120      2475 bp      mRNA      PRI      11-MAR-1998
LOCUS      Homo sapiens mRNA for GS3786, complete cds.
DEFINITION D87120
ACCESSION  D87120.1 GI:1507673
VERSION     GS3786.
KEYWORDS   Homo sapiens cancellous bone tissue; lib:3 end-directed library
SOURCE     osteoblast cell_line:primary-cultured cDNA to mRNA,
           clone_lib:lambda ZAP clone:GS3786.
ORGANISM   Homo sapiens
           Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
           Primates; Catarrhini; Hominoidea; Homo.
REFERENCE  1 (bases 1 to 2475)
AUTHORS   Ohno, I.
TITLE     Direct Submission
JOURNAL   Submitted (14-AUG-1996) to the DDBJ/EMBL/GenBank databases. Ikko
           Ohno, Institute for Molecular and Cellular Biology, Osaka
           University, Molecular Genetics; 1-3 Yamada-oka, Suita, Osaka 565,
           Japan (E-mail:kousaku@imcb.osaka-u.ac.jp, Tel:81-6-879-7992,
           Fax:81-6-877-1923)
REFERENCE  2 (bases 1 to 2475)
AUTHORS   Ohno, I., Hashimoto, J., Takaoka, K., Ochi, T., Okubo, K. and
           Matsubara, K.
TITLE     The cloning of a cDNA for novel genes expressed in human osteoblast
JOURNAL   Unpublished (1996)
FEATURES   Location/Qualifiers
           source
             1..2475
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /cell_line="primary-cultured"
             /cell_type="osteoblast"
             /clone="GS3786"
             /clone_lib="lambda ZAP"
             /tissue_lib="3 end-directed library"
             /tissue_type="cancellous bone"
             168..851
             /codon_start=1
             /product="GS3786"
             /protein_id="BAA13251.1"
             /db_xref="GI:1507674"
             /translation="MRVAGAAKLVAVAVFLITFVVISQVFEIKMDASLGNLFARSAL
             DPAARSTPKPKGKISAKPEKFAKMSGAANVGPKEICLEDNVLMSGVKNVGR
             GINVALAKPKGEVLDTKYEDMGMGDVAPELFELKAIODGTIVLMGTVDGATKLNDR
             ARRLADIGSITNLGFRDNVFCGGKGIKSPFEQHIKNNKDTNKYEGWPEVEM
             EGCIPOKOD"
BASE COUNT 811 a 380 c 527 g 757 t
ORIGIN
Query Match 37.1%; Score 248.2; DB 9; Length 2475;
Best Local Similarity 47.9%; Pred. No. 4.2e-50;
Matches 246; Conservative 94; Mismatches 174; Indels 0; Gaps 0;
QY 151 aartavaartgygnytnathaaarcctgycngcnaaytattytcnttyaarathgtg 210
Db 330 AGATATAAGTGTGGATCTCAAAAGCTGCCCTGGAAGACATTTCTTTAAATATGGCA 389
QY 211 wsgngngcngnaaygtngtngcncnaatgttytygaraygmngnatgathatgwn 270
Db 390 AGTGGAGCACCACGTGGTGGGACCCAAATCTGCTGGAAGATAATCTTTTATGAGT 449
QY 271 ccngtnaaraayaaygtngnmngnnytnaayathgcnnyntngtnaaygncnacngn 330
Db 450 GGTGTTAAGAATAATGTTGGAAGAGGATCAATGTTGCTTGGCAATGGAAACACGGA 509
QY 331 gcngtntngncncaargcnttygavatgtaywsgngngaygtnatgcaytngtnaar 390
Db 510 GAAGATATTAGACACTAAATATTTTTCATGTGGGGAGAGATGTGCACCATTTATTGAG 569
QY 391 ttyytnaarathcncngngngcnnyntngtntngtngcnwnstaygayaycngn 450
Db 570 TTCTGAGGCCATACAGATGGAACAAATAGTTTTTATGGGAACATACCATGATGAGCA 629

QY 451 acnaaragaaygaygarwsmgmnaarytnttywsngaytyngngwnswntaygcnaar 510
Db 630 ACCAACTCAATGATGAGGACGGCGCTATTGCTGATTGGGAGCACAATCTATTACT 689
QY 511 carytngnttymngngaywsntggtnnttyathgngcnaardayytnmngngnaarwsn 570
Db 690 AATCTTGTTTTAGACAGAACTGGGTCTTCTGTGGTGGGAGGCATTAAGACAAAAAGC 749
QY 571 ccnttygarcarttytnaaraaywscngngayacnaayaartaaygargntggcngar 630
Db 750 CCTTTTGAACAGCACAATAAGAACATAAAGATACAAACAATAATGAAGGATGCCTGAA 809
QY 631 ytnytnaragtggargntgyatgccnccnaar 664
Db 810 GTTGTAGAAATGGAGGATGCATCCCCAGAGC 843

RESULT 4
MMU72677/c
LOCUS      Mus musculus EF-7 mRNA, partial cds.
DEFINITION MMU72677
ACCESSION  U72677
VERSION     U72677.1 GI:1872484
KEYWORDS   house mouse.
SOURCE     Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE  1 (bases 1 to 465)
AUTHORS   Fu, X. and Kamps, M.P.
TITLE     E2a-Pbx1 induces aberrant expression of tissue-specific and
           developmentally regulated genes when expressed in NIH 3T3
           fibroblasts
JOURNAL    Mol. Cell. Biol. 17 (3), 1503-1512 (1997)
MEDLINE    97184476
REFERENCE  2 (bases 1 to 465)
AUTHORS   Fu, X. and Kamps, M.P.
TITLE     Direct Submission
JOURNAL    Submitted (27-SEP-1996) Pathology, University of California, San
           Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
FEATURES   Location/Qualifiers
           source
             1..465
             /organism="Mus musculus"
             /strain="BALB/c"
             /db_xref="taxon:10090"
             /cell_line="NIH3T3"
             /cell_type="E2a-Pbx1 induced fibroblasts"
             complement(117..>465)
             /note="similar to human 2.19 protein"
             /codon_start=2
             /product="EF-7"
             /protein_id="AAB51037.1"
             /db_xref="GI:1872485"
             /translation="TGOVMKKDSFDMYSGDPQLLNFLTEIPDSTLVLSVYDDPGTK
             MNDKIKTLFSLNLSYAKQLGFRDSWYFVGAKDLKSKSPYEQLKNNPETNKYDGPWE
             LLEEGCVPRKVM"
BASE COUNT 89 a 129 c 109 g 138 t
ORIGIN
Query Match 28.4%; Score 189.8; DB 12; Length 465;
Best Local Similarity 49.9%; Pred. No. 5.4e-36;
Matches 169; Conservative 74; Mismatches 96; Indels 0; Gaps 0;
QY 325 acngngcngntnyngncncaargcnttygavatgtaywsgngngaygtnatgcaytn 384
Db 464 ACTGCCAGGTGATGAAAAGACAGACTTTGACATGACTCTGGAGACCCCACTCTCTG 405
QY 385 gtnaarttyytnaarathcncngngngcnnyntngtntngtngcnwnstaygayay 444
Db 404 CTGAACCTTCCTACGGAATTCACAGATAGCACCCCTGGTGGTAGCCTCTCTATGATGAC 345
```



```

* 138553 147398: contig of 8846 bp in length
* 147399 147416: gap of unknown length
* 147417 155917: contig of 8501 bp in length
* 155918 155935: gap of unknown length
* 155936 168242: contig of 12307 bp in length
* 168243 168259: gap of unknown length
* 168260 178137: contig of 9878 bp in length.
FEATURES             Location/Qualifiers
     source           1..178137
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /clone="RP11-475023"
BASE COUNT      46731 a 40913 c 41468 g 47960 t 1065 others
ORIGIN
Query Match      13.6%; Score 90.8; DB 55; Length 178137;
Best Local Similarity 52.7%; Pred. No. 2.9e-11;
Matches 69; Conservative 39; Mismatches 23; Indels 0; Gaps 0;
QY 454 aaratgaaygagrwsmgnaarytnttywncdaytngnwnwnstaygcnaarcar 513
Db 109142 AGAATGAACGATGAAGACGAGGAACCTCTCTGACTTGGGGAGTCTCAGCAAAACAA 109083
QY 514 ytingnttymngaywsntgggtnttyatbhgngcnaargayttnmgngnaarwsnccn 573
Db 109082 CTGGGCTCCGGGACAGCTGGGTCTCATAGGACCAAGACCTCAGGGGTAAAGCCCC 109023
QY 574 ttygarcartt 584
||||:||||:|
Db 109022 TTTGAGCAGGT 109012

RESULT 6
HSG6PDGEN 52173 bp DNA PRI 20-AUG-1997
LOCUS H.sapiens G6PD gene for glucose-6-phosphate dehydrogenase.
DEFINITION X55448 229527
ACCESSION X55448.1 GI:450527
VERSION X55448.1
KEYWORDS 2-19 protein; G6PD gene; glucose-6-phosphate dehydrogenase.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 52173)
AUTHORS Zollo,M., Mazzarella,R., Bione,S., Toniolo,D., Schllessinger,D.,
D'Urso,M., and Chen,E.Y.
TITLE Sequence and gene content of the RCP/GCP-G6PD region in human Xq28:
the first 52kb
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 20111)
AUTHORS Chen,E.Y., Cheng,A., Lee,A., Kuang,W.J., Hillier,L., Green,P.,
Schlessinger,D., Ciccodicola,A. and D'Urso,M.
TITLE Sequence of human glucose-6-phosphate dehydrogenase cloned in
plasmids and a yeast artificial chromosome
JOURNAL Genomics 10 (3), 792-800 (1991)
MEDLINE 91365390
REFERENCE 3 (bases 1 to 52173)
AUTHORS Chen,E.Y.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-1994) Ellson Y Chen Dr., Advanced Center for
Genetic Technology, Applied, Biosystems Division, Perkin Elmer
Corp., 850 Lincoln Centre Drive, Foster City, CA, 94404, USA
COMMENT On Jan 26, 1994 this sequence version replaced gi:31540.
FEATURES             Location/Qualifiers
     source           1..52173
                     /organism="Homo sapiens"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="q28"
                     /clone="Cosmid TW1"
                     /clone.lib="pTM by Tom Vulliamy"
misc_difference 1..20111

```

```

/note="1-20111 bp replaces previous submission X55448,
1-20114 bp; 6 corrections made from X55448, including 8850
(delete T), 14575 (add C), 18788 (delete C), 19369 (delete
G), 19389 (delete G), and 19480 (T changes to A); G6PD:
Glucose-6-phosphate dehydrogenase"
/citation=[2]
/label=G6PD
/252..2717
/gene="G6PD"
/citation=[2]
/number=1
/gene="G6PD"
/citation=[2]
/evidence=experimental
/252..18112
/gene="G6PD"
/343..3470
/gene="G6PD"
/citation=[2]
/number=2
/3351..3470,13327..13364,13460..13568,14118..14335,
15007..15165,15343..15468,15834..15927,16375..16561,
16701..16936,17041..17117,17223..17315,17413..17503)
/gene="G6PD"
/citation=[2]
/codon_start=1
/product="glucose-6-phosphate dehydrogenase"
/protein_id="CAA39089.1"
/db_xref="GI:452269"
/db_xref="SWISS-PROT:P11413"
/translation="NAEQVALSRTOVCGLIREELFQGFQDAFHQSDTHIFLIMGAGDLA
KKKIYPTIWLFRDGLLPENTFTMGVARSRLTVADIRKQSEPFPRATPEEKLKLDFF
ARNSVAGYDDAASVQRLSNHDALHILGSOANRLFYALPPTVYEAQTKNHSCMS
QIGWNRKIIVEKFGDRDQSSDRLSNHSLSLFREDQIYRIDHYLGEMQNLMLVRPAN
RFTGPIWNRDNIACVILITKEPFGTEGGGIFDEFGIIRKVNQHLQMLCLVAMEKP
ASTNSDDVRDERKVKVLCISEVQANNVLGYQVNPDGEGEATKGYLDQTPVPGSTT
ATFAAVLVXENERWDGVFILRCGKALNEKAEVRLQFHDVAGIQTIFHQCKRRELVI
RVQNEAVYTKMTKKPGMFENPESELDLTYNRYKNVKLPDAYERLIDVFCGQSOM
HFVRSDELREANRIFTPLLHQIELEKPKPIPIYIGSRGPTADELMKRVGPFQYESTYK
WNPHLK"
/3327..13364
/gene="G6PD"
/citation=[2]
/number=3
13460..13568
/gene="G6PD"
/citation=[2]
/number=4
14118..14335
/gene="G6PD"
/citation=[2]
/number=5
15007..15165
/gene="G6PD"
/number=6
15343..15468
/gene="G6PD"
/number=7
15834..15927
/gene="G6PD"
/number=8
16090
/gene="G6PD"
/note="in the middle of intron 8"
/citation=[1]
/replace="t"
16375..16561
/gene="G6PD"
/number=9
16701..16936
/gene="G6PD"
/number=10

```

```

exon      17041..17117
/gene="G6PD"
/number=11
variation 17210
/gene="G6PD"
/note="near the end of intron 11"
/citation=[1]
/replace="c"
17223..17315
/gene="G6PD"
/number=12
exon      17413..18112
/gene="G6PD"
/number=13
variation 17860
/gene="G6PD"
/note="exon13, 3'UTR"
/citation=[1]
/replace="g"
18094..18099
/gene="G6PD"
18112
/gene="G6PD"
/citation=[2]
evidence-experimental
19058
/note="downstream G6PD intergenic region"
/citation=[1]
/replace="t"
33000..33289
/gene="2-19 gene"
/citation=[1]
/label=exon1
/number=1
prim_transcript 33000..42862
/gene="2-19"
/citation=[1]
evidence-experimental
33000..33289
/gene="2-19 gene"
33000..42862
/gene="2-19"
Join(33277..33289,36113..36198)
/gene="2-19"
Join(33277..33289,36113..36226,37166..37189,40439..40562,
40689..40747,41175..41225,41546..41630,41707..41833,
42131..42226)
/gene="2-19"
/citation=[1]
/codon_start=1
/label=protein2-19
evidence-experimental
/product="2-19 protein"
/protein_id="CAA39090.1"
/db_xref="GI:452270"
/db_xref="SWISS-PROT:P98173"
/translacion="MRLAGPLRIVLVSVGVTVIIVSVLLGGPGSGFPRIQQLFTSP
ESSVTAAPRARKYKCGLPQPCPEEHLAFRVDSGAANVIGPKICLEDKMLMSSVKONVG
RGLNIALVNGVSGELTEARAFDMWAGVDNLLKFIPLHEGTLVFAVSYDDPATKME
ETRLFLSELGSRKELAFRDSWVFGAKGVQNKSPFEQHVKNKSHKSYEGCPENALE
MECCIPRRSTAS"
36113..36226
/gene="2-19"
/citation=[1]
/label=exon2
/number=2
37166..37189
/gene="2-19"
/citation=[1]
/label=exon3
/number=3
40439..40562
/gene="2-19"

```

```

/citation=[1]
/label=exon4
/number=4
40689..40747
/gene="2-19"
/citation=[1]
/label=exon5
/number=5
41175..41225
/gene="2-19"
/citation=[1]
/label=exon6
/number=6
41546..41630
/gene="2-19"
/citation=[1]

Query Match      11.4%; Score 76.2; DB 9; Length 52173;
Best Local Similarity 45.4%; Pred. No. 8.1e-08;
Matches 64; Conservative 37; Mismatches 40; Indels 0; Gaps 0;

QY 444 yccngnacnaaratgaaygaygarwsmngnaarytnttywsngayytnngnwnsnta 503
      ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 41695 CCATGCTGCAGGATGATGAAGAGAGACCAGAAAGCTCTTCAGTGAGCTGGCAGCAGGAA 41754

QY 504 ygcnaarcarytngnttymgngaywsntgtgnttyathgngcnaargayytnmngng 563
      :|| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 41755 CGCCAAGGAGCTGGCCTTCGGGACAGCTGGTGTTGTCGGGGCCACAGGGTGTGCAGAA 41814

QY 564 naarwsnccnttygarcartt 584
      |||: || |||: |||: |
Db 41815 CRAAGAGCCCTTTGAGCAGGT 41835

RESULT 7
HUMFLNG6PD/c HUMFLNG6PD 219447 bp DNA PRI 17-MAY-1999
LOCUS Homo sapiens chromosome X region from filamin (FLN) gene to
DEFINITION glucose-6-phosphate dehydrogenase (G6PD) gene, complete cds's.
ACCESSION L44140
VERSION L44140.1 GI:1203968
KEYWORDS 1A gene; 2-19 gene; ABP-280 gene; DNLI1L gene; DNase I-like protein
gene; EMD gene; FLN gene; G4.5 gene; G4.8 gene; G6PD gene; GDI
gene; Gdx gene; P3 gene; QM gene; STA gene; XAP-1 gene; XAP-2 gene;
XAP-4 gene; XAP-5 gene; XAP-7 gene; actin-binding protein; emerlin;
emerly-dreyfuss syndrome; filamin; glucose-6-phosphate
dehydrogenase.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 219447)
AUTHORS Chen,E.Y., Zollo,M., Mazzarella,R.A., Ciccodicola,A., Chen,C.-N.,
Zuo,L., Helner,C., Burrough,F.W., Ripetto,M., Schlessinger,D. and
D'Urso,M.
TITLE Long-range sequence analysis in Xq28: thirteen known and six
candidate genes in 219.4 kb of high GC DNA between the RCP/GCP and
G6PD loci
HUM. MOL. GENET. 5 (5), 659-668 (1996)
JOURNAL 96311563
MEDLINE
COMMENT Submitted by:
Ellison Chen,
Advanced Center for Genetic Technology,
Applied Biosystems Division of Perlin Elmer Corp., 850 Lincoln
Center Drive,
Foster City, CA 9404 USA
and
David Schlessinger,
Department of Molecular Microbiology and Center for Genetics in
Medicine
Washington University School of Medicine,
St. Louis MO 63110 USA
e-mail: ellison@enseq.apldbio.com and david@genetics.wustl.edu

```

FEATURES
source

Note: Gene predictions were accomplished with runs of Grail versions 1.1 and 1.2, coupled with fasta and blastx comparisons to genbank & non-redundant peptide libraries. Repeat analysis was accomplished via censor.

Location/Qualifiers

repeat_unit	1. .219447 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="X" /map="X"	repeat_unit	9486. 9773 /rpt_family="Alu-J" /evidence=experimental 9996. .10132
repeat_unit	4. .223 /rpt_family="Alu-Sp or Alu-Sq" /evidence=experimental 226. .503	repeat_unit	/rpt_family="Alu-Sp" /evidence=experimental 10159. 10255
repeat_unit	/rpt_family="Alu-J or Alu-S" /evidence=experimental 641. .931	repeat_unit	/rpt_family="Alu-Sq" /evidence=experimental 10271. .10367
repeat_unit	/rpt_family="Alu-J or Alu-S" /evidence=experimental 1023. .1313	repeat_unit	/rpt_family="Alu-J" /evidence=experimental complement(10455. .10739)
repeat_unit	/rpt_family="Alu-Sx" /evidence=experimental complement(1446. .1536)	repeat_unit	/rpt_family="Alu-Sx" /evidence=experimental complement(10740. .10860)
repeat_unit	/rpt_family="Alu-J" /evidence=experimental 3095. .3426	repeat_unit	/rpt_family="Alu-J or Alu-S" /evidence=experimental complement(10866. .11156)
repeat_region	/rpt_family="CpG Island" /evidence=experimental complement(3663. .3952)	repeat_unit	/rpt_family="Alu-Sx" /evidence=experimental 11223. .11501
repeat_unit	/rpt_family="Alu-Sb0 or Alu-Sb1" /evidence=experimental 4912. .5046	repeat_unit	/rpt_family="Alu-J" /evidence=experimental 11638. .13564
repeat_unit	/rpt_family="Alu-J or Alu-S" /evidence=experimental 5047. .5333	repeat_unit	/rpt_family="CpG Island" /evidence=experimental complement(11884. .12163)
repeat_unit	/rpt_family="Alu-Sb0 or Alu-Sb1" /evidence=experimental 5346. .5628	repeat_unit	/rpt_family="Alu-Sb0 or Alu-Sb1" /evidence=experimental complement(12195. .12482)
repeat_unit	/rpt_family="Alu-J or Alu-S" /evidence=experimental 5632. .5805	repeat_unit	/rpt_family="Alu-J or Alu-S" /evidence=experimental complement(12668. .12945)
repeat_unit	/rpt_family="Alu-Sq" /evidence=experimental 6363. .17698	repeat_unit	/rpt_family="Alu-J or Alu-S" /evidence=experimental complement(18360. .18648)
repeat_unit	/rpt_family="11kb repeat1" /evidence=experimental 6369. .6633	exon	/rpt_family="Alu-Sb0" /evidence=experimental complement(18984. .19488)
repeat_unit	/rpt_family="Alu-Sx" /evidence=experimental complement(7121. .7409)	gene	/gene="FLN" /number=48 /evidence=experimental complement(18984. .45006)
repeat_unit	/rpt_family="Alu-Sp" /evidence=experimental complement(7358. .7432)	CDS	complement(join(19301. .19488,19814. .20017,20101. .20319,20483. .20659,21361. .21493,22033. .22148,22336. .22473,22633. .22899,23005. .23127,23224. .23376,23453. .23656,23748. .23909,24006. .24179,24367. .24495,24603. .24743,24833. .24935,25068. .25163,25277. .25524,27701. .27724,27884. .28073,28649. .28805,28895. .29018,29434. .29604,29696. .29856,29934. .30096,30182. .30355,30440. .31037,31758. .32020,32120. .32237,32429. .32598,32692. .32782,32868. .33028,33111. .33234,34472. .34615,34709. .34822,34976. .35169,35271. .35407,35586. .35709,35798. .35935,36473. .36673,36757. .36919,37011. .37088,37181. .37399,37846. .37993,38090. .38187,38291. .38539,41319. .41691))
repeat_unit	/rpt_family="Alu-Sb0 or Alu-Sb1" /evidence=experimental complement(7426. .7701)		/gene="FLN"
repeat_unit	/rpt_family="Alu-J" /evidence=experimental 8306. .8592		/codon_start=1 /evidence=experimental
repeat_unit	/rpt_family="Alu-Sx" /evidence=experimental complement(8863. .9029)		/product="filamin"
repeat_unit	/rpt_family="Alu-J or Alu-S" /evidence=experimental complement(9031. .9317)		/protein_id="AAA92644.1"
repeat_unit	/rpt_family="Alu-Sq" /evidence=experimental complement(9318. .9442)		/db_xref="GI:1203969"
repeat_unit	/rpt_family="Alu-J or Alu-S" /evidence=experimental		/translation="MSSSHSRAGQSAAGAAGGGVDTDAEMPATEKDLAEDAPWKKI QONTFRWNEHLKCVSKRIANLQDLSGLLIALLEVLVSQKMHKHKHQRPTFRM OLENVSVALEFLDRESIKLSIDSKAIVDGNLKLILGLIWLILHYSISMPWDEED EAKKOTPKORLLGIONKLPOLPITNFSRWQSGRALGVDSACGLCPMDWSDA SKPVTNAREAMQADDNLGIPQVITPEIIVPDVDEHSVMTYLSQFPKALKKCAPLR PKLKKARAGYGGIEPTGNKKRAEFTVRSAGOGVELVYVDEDPAGHOEEAKVTA NDKNRTFSVMYVPEVTGTHKVTILFAGQIAKSPFEYVDKSGQDASKYTAQGPGL PSGNIANKTTFEITAGAGTGEVVIQDPMQKGTVEPQLEARGDSTYRCYQPTM EGVTHVHTFAGVPIPRSPYTVTVGQACNPSACRAVGRGLQPKGVRYKETADEFVTK GAGSGELKVTYKGPKEERYKQKDLGDGVYGFYYPMVPGTYIVTIWGGQNGIRSPF

[illegible]

Search completed: May 31, 2000, 18:42:56
Job time: 5915 sec

[illegible]

RESULT	15	AC006745	HTG	23-FEB-1999
LOCUS		AC006745	DNA	23-FEB-1999
DEFINITION		Caenorhabditis elegans clone Y38F2, *** SEQUENCING IN PROGRESS ***, 5 unordered pieces.		
ACCESSION		AC006745		
VERSION		AC006745.1	GI:4263193	
KEYWORDS		HTG; HTGS_PHASE1.		
SOURCE		Caenorhabditis elegans.		
ORGANISM		Caenorhabditis elegans.		
REFERENCE		Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida; Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis		
AUTHORS		1 (bases 1 to 249287)		
TITLE		Waterston,R.H.		
JOURNAL		The sequence of Caenorhabditis elegans clone		
REFERENCE		Unpublished		
AUTHORS		2 (bases 1 to 249287)		
TITLE		Waterston,R.H.		
JOURNAL		Direct Submission		
COMMENT		Submitted (23-FEB-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis MO 63108, USA		
		* NOTE: This is a 'working draft' sequence. It currently		
		* consists of 5 contigs. The true order of the pieces		
		* is not known and their order in this sequence record is		
		* arbitrary. Gaps between the contigs are represented as		
		* runs of N, but the exact sizes of the gaps are unknown.		
		* This record will be updated with the finished sequence		
		* as soon as it is available and the accession number will		
		* be preserved.		
		1 147084: contig of 147084 bp in length		
		* 147085 147093: gap of unknown length		
		* 147094 187319: contig of 40226 bp in length		
		* 187320 187328: gap of unknown length		
		* 187329 189480: contig of 2152 bp in length		
		* 189481 189489: gap of unknown length		
		* 189490 245636: contig of 56147 bp in length		
		* 245637 245644: gap of unknown length		
		* 245645 249287: contig of 3643 bp in length.		

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: May 31, 2000, 17:04:21 ; Search time 1292.59 seconds
(without alignments)
-918.913 Million cell updates/sec
Title: US-09-186-342-1
Perfect score: 1221
Sequence: 1 cgccgggaggtgcatct.....aaatttttttgcgtga 1221

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 882769 seqs, -486395729 residues
Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0
Maximum DB seq length: 1000000
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba1.*
2: gb_ba2.*
3: gb_om.*
4: gb_ov.*
5: gb_pat.*
6: gb_ph.*
7: gb_p11.*
8: gb_p12.*
9: gb_p1.*
10: gb_p12.*
11: gb_p13.*
12: gb_ro.*
13: gb_ro.*
14: gb_sy.*
15: gb_un.*
16: gb_v1.*
17: em_fun.*
18: em_hum1.*
19: em_hum2.*
20: em_in.*
21: em_om.*
22: em_or.*
23: em_ov.*
24: em_pat.*
25: em_ph.*
26: em_pi.*
27: em_ro.*
28: em_ro.*
29: em_sy.*
30: em_un.*
31: em_v1.*
32: gb_htg1.*
33: gb_htg2.*
34: gb_in1.*
35: gb_in2.*
36: em_ba1.*
37: em_ba2.*
38: em_bum3.*
39: em_bum4.*
40: gb_p14.*
41: gb_htg3.*
42: gb_htg4.*
43: gb_htg5.*
44: gb_htg6.*

45: gb_htg7.*
46: em_htg1.*
47: em_htg2.*
48: em_htg3.*
49: em_hum5.*
50: gb_p13.*
51: gb_p15.*
52: gb_htg8.*
53: gb_htg9.*
54: gb_htg10.*
55: gb_htg11.*
56: gb_htg12.*
57: gb_htg13.*
58: gb_htg14.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	515.6	42.2	548	5	I95745	I95745 Sequence 6
C 2	404.2	33.1	178137	55	AC023346	AC023346 Homo sapi
3	247.8	20.3	1606	9	HS219MRNA	X87193 H.sapiens m
C 4	229.4	18.8	465	12	MMU72677	U72677 Mus musculu
5	216.4	17.7	2475	9	D87120	D87120 Homo sapien
6	70	5.7	52173	9	HSG6PDGEN	X55448 H.sapiens G
C 7	56.6	4.6	132319	40	AC006364	L44140 Homo sapien
8	46.6	3.8	7218	5	I66494	AC006364 Homo sapi
C 9	43.4	3.6	35052	57	AC016137	I66494 Sequence 14
C 10	43	3.5	1302	51	AF090946	AC016137 Homo sapi
11	43	3.5	1302	51	AF090946	AF090946 Homo sapi
C 12	43	3.5	173823	33	HSJ963E22	AL096828 Homo sapi
C 13	41.4	3.4	123645	33	HSJ697K14	AL121829 Homo sapi
C 14	41.4	3.4	185436	33	HSDJ663D7	AL121907 Homo sapi
15	39.8	3.3	10969	34	LMFL4768	AL132763 Leishmani
16	39.6	3.2	112626	41	AC008616	AC008616 Homo sapi
17	39.2	3.2	2440	5	A20492	AC008616 Homo sapi
C 18	39.2	3.2	2440	5	A20493	A20492 cDNA insert
19	39.2	3.2	2458	5	A20495	A20493 cDNA insert
C 20	39.2	3.2	2458	5	A20495	A20495 vector pT7/
C 21	39.2	3.2	3086	5	A20498	A20496 vector pT7/
22	39.2	3.2	3086	9	HSIGREX	A20498 RFX gene. 7
23	38.4	3.1	178896	40	AC008041	X58964 H.sapiens g
C 24	38.4	3.1	283438	57	AC012465	AC008041 Homo sapi
C 25	38.2	3.1	173561	32	AP001141	AC012465 Homo sapi
26	38	3.1	1351	12	AF022079	AF001141 Homo sapi
27	38	3.1	3233	12	AF027181	AF022079 Mus muscu
28	38	3.1	4505	12	AF016252	AF027181 Rattus no
29	38	3.1	6803	11	HSU86758	AF016252 Rattus no
C 30	37.8	3.1	2976	2	AF226283	U86758 Human netri
31	37.4	3.1	55590	56	AC011531	AF226283 Xanthomon
32	37.2	3.0	2454	10	HSSGI	AC011531 Homo sapi
C 33	37.2	3.0	120940	55	AC018939	Y00064 Human mRNA
34	37.2	3.0	139352	10	HS967N21	AC018939 Homo sapi
C 35	37	3.0	2282	9	HUMBSLAA	AL035461 Human DNA
36	37	3.0	5200	9	HUMBSNF28	M31731 Human chron
37	37	3.0	5247	10	HSU29175	D26156 Human mRNA
38	37	3.0	23550	1	SCE39	U29175 Human trans
39	36.8	3.0	160218	45	AC010335	AL049573 Streptomy
40	36.4	3.0	74079	45	AC021480	AC010335 Homo sapi
41	36.4	3.0	163253	43	AC011978	AC021480 Mus muscu
C 42	36	2.9	59438	42	AC015710	AC011978 Homo sapi
C 43	36	2.9	216347	33	AL138885	AC015710 Homo sapi
44	35.8	2.9	380	13	G54682	AL138885 Homo sapi
45	35.8	2.9	380	13	G60062	G54682 AffySTS0000

ALIGNMENTS

```
RESULT 1
195745
LOCUS 195745 548 bp DNA PAT 01-DEC-1998
DEFINITION Sequence 6 from patent US 5733748.
ACCESSION 195745
VERSION 195745.1 GI:3940215
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE
1 (bases 1 to 548)
AUTHORS Yu,G. and Rosen,C.
TITLE Colon specific genes and proteins
JOURNAL Patent: US 5733748-A 6 31-MAR-1998;
FEATURES
Location/Qualifiers
1..548
source
/organism="unknown"
BASE COUNT 137 a 128 c 173 g 107 t 3 others
ORIGIN

Query Match
Best Local Similarity 42.2%; Score 515.6; DB 5; Length 548;
Matches 540; Conservative 1; Mismatches 7; Indels 2; Gaps 2;

QY 486 atgagctctgtgaaaaacaatgtgggagagggcctaacaatccctctgtgtaatggaacc 545
Db 1 ATGAGTCTCTGTGAAAAACAATGTGGGCAGAGGGCCTAAACATCCCTGTGTGTAATGGAACC 60
QY 546 acgggagctgtgctgggacagaagcatttgacatgtactctctggagatgttatgcacct 605
Db 61 ACGGGAGCTGTCTGGGACAGACGATTTGACATGTACTCTGGAGATGTTATGCACCTA 120
QY 606 gtgaattccttaagaataatccgggggggtgactgggtgctgtgctcctctcaacagat 665
Db 121 GTGAATTCCTTAAAGAAATTCGGGGGGGTGCACGTGTGTGTGCTCTCCCTACGACGAT 180
QY 666 ccagggacacaaatgaacgatgaagcaggagaaactctctctgaacttggggagttcctac 725
Db 181 CCAGGACCAAAATGAACGATGAAGCAGGAAACTCTCTGACTTGGGGAGTCTCTTAC 240
QY 726 gcaaacacactggcttcgggacagctgggtcttctataggacccaagaacctcaggggt 785
Db 241 GCAAACACACTGGGCTTCGGGACAGCTGGGTCTTCATAGGAGCCCAAGACCTCAGGGGT 300
QY 786 aaaaagccctttgagcagttctttaaagaacagcccgacacacaaacaaacacagggatgg 845
Db 301 AAAAGCCCTTTGAGCAGTCTTAAAGAACAGCCCGACACACAAACAAATACGAGGGATGG 360
QY 846 ccagagctgtctgagatgagagggtgctgacccccccgaagccattttaggggtgtgtgac 905
Db 361 CCAGAGCTGTCTGGAGATGGAGGGTGTGATCCCCCGAAGCCATTTTAGGGTGGCTGTGGC 420
QY 906 ttctctcagccaggggctgaagaagctcctgctgaacttaggtcagagcccgccggcag 965
Db 421 TCTTCTCAGCCAGGGGCTGACAGAGYCTCTGCTGATAGGAGTCANAGCCCGGCA - 480
QY 966 gggctgaagagagagagagaggggtgtgctgctggaaggtgtctcaggtcctctgacgctg 1025
Db 480 GGCTGNAGGAGGAGGAGGAGGAGGGGTGTGCTGTGGAAGGTGTCTGACAGG-CCTTGCACGCTG 538
QY 1026 tgtcgcgcct 1035
Db 539 TGTGCGCCT 548

RESULT 2
AC023346/c
LOCUS 178137 bp DNA HTG 11-FEB-2000
DEFINITION Homo sapiens clone RP11-475023, *** SEQUENCING IN PROGRESS ***, 53
ACCESSION AC023346
unordered pieces.
```

```
AC023346.1 GI:5960447
HTG; HTGS_PHASE1.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 178137)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
2 (bases 1 to 178137)
Waterston,R.H.
Direct Submission
Submitted (11-FEB-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
Center project name: H.NH0475023.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 53 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1 1212: contig of 1212 bp in length
* 1213 1230: gap of unknown length
* 1231 2338: contig of 1108 bp in length
* 2339 2356: gap of unknown length
* 2357 3517: contig of 1161 bp in length
* 3518 3535: gap of unknown length
* 3536 4850: contig of 1315 bp in length
* 4851 4868: gap of unknown length
* 4869 6528: contig of 1660 bp in length
* 6529 6546: gap of unknown length
* 6547 7838: contig of 1292 bp in length
* 7839 7856: gap of unknown length
* 7857 9445: contig of 1589 bp in length
* 9446 9463: gap of unknown length
* 9464 11266: contig of 1803 bp in length
* 11267 11284: gap of unknown length
* 11285 12732: contig of 1448 bp in length
* 12733 12750: gap of unknown length
* 12751 14136: contig of 1386 bp in length
* 14137 14154: gap of unknown length
* 14155 15188: contig of 1034 bp in length
* 15189 15206: gap of unknown length
* 15207 17203: contig of 1997 bp in length
* 17204 17221: gap of unknown length
* 17222 18766: contig of 1545 bp in length
* 18767 18784: gap of unknown length
* 18785 20759: contig of 1975 bp in length
* 20760 20777: gap of unknown length
* 20778 22676: contig of 1899 bp in length
* 22677 22694: gap of unknown length
* 22695 24767: contig of 2073 bp in length
* 24768 24785: gap of unknown length
* 24786 26285: contig of 1500 bp in length
* 26286 26303: gap of unknown length
* 26304 27755: contig of 1452 bp in length
* 27756 27773: gap of unknown length
* 27774 29754: contig of 1981 bp in length
* 29755 29772: gap of unknown length
* 29773 32098: contig of 2326 bp in length
* 32099 32116: gap of unknown length
* 32117 34385: contig of 2269 bp in length
* 34386 34403: gap of unknown length
* 34404 36992: contig of 2589 bp in length
* 36993 37010: gap of unknown length
* 37011 39118: contig of 2108 bp in length
* 39119 39136: gap of unknown length
* 39137 40831: contig of 1695 bp in length
* 40832 40849: gap of unknown length
```

[illegible]

ETKLFSELSRRAKELAFRDSWVFVGAQVQNKSPFEQHVKNKSHNKEGCEPALE

MEGIPRRSTAS

1588..1593

polya_signal 301 a 521 c 503 g 281 t

BASE COUNT

ORIGIN

Query Match 20.3%; Score 247.8; DB 9; Length 1606;
Best Local Similarity 66.2%; Pred. No. 3.1e-56;
Matches 357; Conservative 0; Mismatches 182; Indels 0; Gaps 0;

QY 344 ctgcgccaccaagagatccagggtataaaagtacaaagtgtggtccctacataagccctgcc 403
Db 415 ctgcgtgctgcagccgacgagcagcagcagcagcagcagcagcagcagcagcagcagc 474
QY 404 agcccaactactctgtggtttataaatctgcagtggtggcgccgcaacgctggtggccctactat 463
Db 475 TGAGGAGCACCTGGCTTCCGCGTGTGCAGCGGGCGCGCAACGTCATTGGGCCCAAGAT 534
QY 464 gtgtttgaagaccgcagcatgatcgtctgtgtaaaacaatagtgggcagagcctataa 523
Db 535 CTGCTCTGAGGACAAAGATGCTGATGAGCAGCGTCAAGGACAACTGGGGCGCGGCTGAA 594
QY 524 catcgcctcgtggaatgaacacagcagcagcagcagcagcagcagcagcagcagcagc 583
Db 595 CATCGCCTGTGTGAACGGGGTCAGCGCGAGCTCATCGAGCGCGCGGCTTGACATGTG 654
QY 584 ctctggagatgtatgcacctagtgaattcctttaaagaaattccgggggtgcactggt 643
Db 655 GCGCCGAGATGTCAACGACCTGTTGAAGTTTATTCGGCCATGACGACGACCCCTGGT 714
QY 644 gctggtggcctctacacgacgtccagggaccacaaatgaacgataaagcaggaacctct 703
Db 715 GTTGTGGCATCTTACGAGCGCCACGACCAAGATGAATGAAGAGACAGAAAGCTCTT 774
QY 704 ctctgactgtgggagttcctactcgaacaaactgggttccggagacagctgggtcttcat 763
Db 775 CAGTAGCTGGGACGACGAGCAAGCGCTGGCGCTTCCGGACACAGCTGGGTGTTGT 834
QY 764 agggagcaaacgacctcagggttaaaagccctttgagcagttcttaagcaacagccaga 823
Db 835 CGGGCCRAAGGTGTGCAGACACAGAGCCCTTTGAGCAGCAGCAGTGAAGACAGTAAGCA 894
QY 824 cacaacaaataacagagagatgcccagagctgctggagatggagggctgcagccccga 882
Db 895 CAGCAACAAGTAGCAAGGCTGCCCGAGGCGCTGGAGCTGGAGAGCTGTATCCCGCGGA 953

RESULT 4
LOCUS MMU72677 465 bp mRNA ROD 29-MAR-1997
DEFINITION Mus musculus EF-7 mRNA, partial cds.
ACCESSION U72677
VERSION U72677.1 GI:1872484
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 465)
AUTHORS Fu,X. and Kamps,M.P.
TITLE E2a-Pbx1 induces aberrant expression of tissue-specific and developmentally regulated genes when expressed in NIH 3T3 fibroblasts
JOURNAL Mol. Cell. Biol. 17 (3), 1503-1512 (1997)
MEDLINE 97184476
REFERENCE 2 (bases 1 to 465)
AUTHORS Fu,X. and Kamps,M.P.
TITLE Direct Submission
JOURNAL Submitted (27-SEP-1996) Pathology, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093, USA
FEATURES Location/Qualifiers
source 1..465

/organism="Mus musculus"
/strain="BALB/c"
/db_xref="taxon:10090"
/cell_line="NIH3T3"
/cell_type="E2a-Pbx1 induced fibroblasts"
/complement(117..>465)
/note="similar to human 2.19 protein"
/codon_start=2
/product="EF-7"
/protein_id="AAB51037.1"
/db_xref="GI:1872485"
/translation="TGVVKKDSFDMYSGDPQLLLNFIETEPDSTLVLVASYDDPGTK
MNKIKTLFNLGSSYAKQLGFRDWSVVFVGAQVQNKSPFEQHVKNKSHNKEGCEPALE
LLEEGCVPKRVN"
BASE COUNT 89 a 129 c 109 g 138 t
ORIGIN

Query Match 18.8%; Score 229.4; DB 12; Length 465;
Best Local Similarity 77.4%; Pred. No. 2.9e-51;
Matches 278; Conservative 0; Mismatches 81; Indels 0; Gaps 0;
QY 545 acgggagctgtcgtgggacagaagcatttgacatgtactctgagatgttatgcaccta 605
Db 464 ACTGCCAGGTGATGAAAAGGACAGCTTTGACATGTACTCTGGAGACCCCAACTCCTG 405
QY 606 gtgaattccttaagaattccgggggtgctgctggtggtggtggtggtggtggtggtggt 665
Db 404 CTGAATCTCTTACGGAATTCACAGATAGCACCTGTGTGTAGCTCTCTATGATGAC 345
QY 666 ccagggacaaaatgaacgataaagcaggaactctctctgacttggggaggttccctac 725
Db 344 CCAGGACCAAAATGAAGCAAAATCAAGACGCTTTTCTCCAACCTGGGAAGTCTCTAC 285
QY 726 gcaaaacaaactgggtccgggacagctgggttcttcatagagccaaagacctcagggt 785
Db 284 GCCAAGCAGCTGGGCTTCCGGGACAGCTGGGTCTTTGTAGGTCTAAAGACCTCAAGAGT 225
QY 786 aaaaagccctttagcagcttcttaaaagacagccagacacacacacacacacacacac 845
Db 224 AAAAGCCCATATGAGCAGTCTTTAAAGAACACACCCAGAACAAACAAATATGACGATG 165
QY 846 ccagagctgctggagatgagggctgcatgcccccgaagccatttaggggtgctgtg 904
Db 164 CCGGAGTTGCTAGAGCTCGAGGCTGTGCGCCAGGAAGTAATGTAGGTGGCCTAGG 106

RESULT 5
LOCUS D87120 2475 bp mRNA PRI 11-MAR-1998
DEFINITION Homo sapiens mRNA for GS3786, complete cds.
ACCESSION D87120
VERSION D87120.1 GI:1507673
KEYWORDS GS3786.
SOURCE Homo sapiens cancellous bone tissue lib:3 end-directed library osteoblast cell_line:primary-cultured CDNA to mRNA, clone_lib:lambda ZAP clone:GS3786.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2475)
AUTHORS Ohno,I.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-1996) to the DDBJ/EMBL/GenBank databases. Ikko Ohno, Institute for Molecular and Cellular Biology, Osaka University, Molecular Genetics, 1-3 Yamada-oka, Suita, Osaka 565, Japan (E-mail:kousaku@imcb.osaka-u.ac.jp, Tel:81-6-879-7992, Fax:81-6-877-1922)
REFERENCE 2 (bases 1 to 2475)
AUTHORS Ohno,I., Hashimoto,J., Takaoka,K., Ochi,T., Okubo,K. and Matsubara,K.
TITLE The cloning of a cDNA for novel genes expressed in human osteoblast
JOURNAL Unpublished (1996)


```

exon      13327..13364
/gene="G6PD"
/citation=[2]
/number=3
13460..13568
/gene="G6PD"
/citation=[2]
/number=4
14118..14335
/gene="G6PD"
/citation=[2]
/number=5
15007..15165
/gene="G6PD"
/number=6
15343..15468
/gene="G6PD"
/number=7
15834..15927
/gene="G6PD"
/number=8
16090
/gene="G6PD"
/notes="in the middle of intron 8"
/citation=[1]
/replace="t"
16375..16561
/gene="G6PD"
/number=9
16701..16936
/gene="G6PD"
/number=10
17041..17117
/gene="G6PD"
/number=11
17210
/gene="G6PD"
/notes="near the end of intron 11"
/citation=[1]
/replace="c"
17223..17315
/gene="G6PD"
/number=12
17413..18112
/gene="G6PD"
/number=13
17860
/gene="G6PD"
/notes="exon13, 3'UTR"
/citation=[1]
/replace="g"
18094..18099
/gene="G6PD"
18112
/gene="G6PD"
/evidence=experimental
19058
/notes="downstream G6PD intergenic region"
/citation=[1]
/replace="t"
33000..33289
/gene="2-19 gene"
/citation=[1]
/label=exon1
/number=1
33000..42862
/gene="2-19"
/evidence=experimental
33000..33289
/gene="2-19 gene"
33000..42862

```

```

slg_peptide
CDS
/gene="2-19"
join(33277..33289,36113..36198)
/gene="2-19"
join(33277..33289,36113..36226,37166..37189,40439..40562,
40689..40747,41175..41225,41546..41630,41707..41833,
42131..42226)
/gene="2-19"
/citation=[1]
/codon_start=1
/label=protein2-19
/evidence=experimental
/product="2-19 protein"
/protein_id="CAA39090.1"
/db_xref="GI:452270"
/db_xref="SWISS-PROT:P98173"
/translation="MRLAGPLRIYLVVSVGVTVIVVSTLLGGPGSGFPRIQLETSP
ESSVTAAPRARKYKGLPQPCPEHFLAFRVVSGAANVIGKICLEDKMLMSSYKDNVG
RGNTALVNGVSGELIEARFDMWAGVDNLLKPIRPLHEGTLVVFVSYDDPATKME
ETKLFSELGSSNAKELAFRDSWVFVGAKGVQNKSPFEOHVKNKSHKYECPLEALE
MEGCIPRRSTAS"
36113..36226
/gene="2-19"
/citation=[1]
/label=exon2
/number=2
37166..37189
/gene="2-19"
/citation=[1]
/label=exon3
/number=3
40439..40562
/gene="2-19"
/citation=[1]
/label=exon4
/number=4
40689..40747
/gene="2-19"
/citation=[1]
/label=exon5
/number=5
41175..41225
/gene="2-19"
/citation=[1]
/label=exon6
/number=6
41546..41630
/gene="2-19"
/citation=[1]

```

```

Query Match      5.7%; Score 70; DB 9; Length 52173;
Best Local Similarity 68.3%; Pred. No. 2.3e-08;
Matches 97; Conservative 0; Mismatches 45; Indels 0; Gaps 0;
QY 666 ccagggaacaaatgaacatgaagcaggaactctctctgacttggggagttctac 725
||| | || ||||| || ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41696 CCATGCTGCAGATGAATGAAGAGACCAGAAAGCTCTTCAGTGAGCTGGCAGCAAC 41755
QY 726 gcaaacacactggctccgggacagctggtcttcataaggagccaaagacctcagggt 785
||| | || ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41756 GCCAAGGAGCTGCCCTCCGGGACAGCTGGGTGTTGTGCGGGCCCAAGGGTGTGCAGAAC 41815
QY 786 aaagagccctttgagcagttct 807
|| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41816 AAGAGCCCCCTTGAGCAGGTAT 41837

```

```

RESULT 7
HUMFLNG6PD/c
LOCUS HUMFLNG6PD 219447 bp DNA PRI 17-MAY-1999
DEFINITION Homo sapiens chromosome X region from filament (FLN) gene to
glucose-6-phosphate dehydrogenase (G6PD) gene, complete cds's.
ACCESSION U44140
VERSION U44140.1 GI:1203968

```

KEYWORDS	1A gene; 2.19 gene; ABP-280 gene; DNLL gene; DNase I-like protein gene; EMD gene; FLN gene; G4.5 gene; G4.8 gene; G6PD gene; GDI gene; Gdx gene; P3 gene; QM gene; STA gene; XAP-1 gene; XAP-2 gene; XAP-4 gene; XAP-5 gene; XAP-7 gene; actin-binding protein; emerlin; emery-dreyfuss syndrome; filamin; glucose-6-phosphate dehydrogenase.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	Chen,E.Y., Zollo,M., Mazzarella,R.A., Ciccodicola,A., Chen,C.-N., Zuo,L., Heiner,C., Burrough,F.W., Ripetto,M., Schlössinger,D. and D'Urso,M.
TITLE	Long-range sequence analysis in Xq28: thirteen known and six candidate genes in 219.4 kb of high GC DNA between the RCP/GCP and G6PD loci
JOURNAL	Hum. Mol. Genet. 5 (5), 659-668 (1996)
MEDLINE	96311563
COMMENT	Submitted by: Ellson Chen, Advanced Center for Genetic Technology, Applied Biosystems Division of Perlin Elmer Corp., 850 Lincoln Center Drive, Foster City, CA 94404 USA and David Schlössinger, Department of Molecular Microbiology and Center for Genetics in Medicine Washington University School of Medicine, St. Louis MO 63110 USA e-mail: ellson@genseq.aplbbio.com and david@genetics.wustl.edu Note: Gene predictions were accomplished with runs of Grail versions 1.1 and 1.2, coupled with fasta and blastx comparisons to genbank & non-redundant peptide libraries. Repeat analysis was accomplished via a censor.
FEATURES	Location/Qualifiers 1..219447 /organism="Homo sapiens" /db_xref="taxon:9606" /chromosome="X" /map="X"
repeat_unit	4..223 /rpt_family="Alu-Sp or Alu-Sq" /evidence=experimental
repeat_unit	226..503 /rpt_family="Alu-J or Alu-S" /evidence=experimental
repeat_unit	641..931 /rpt_family="Alu-Sq" /evidence=experimental
repeat_unit	1023..1313 /rpt_family="Alu-J or Alu-S" /evidence=experimental
repeat_unit	1023..1313 /rpt_family="Alu-Sx" /evidence=experimental
repeat_unit	complement(1446..1536) /rpt_family="Alu-J" /evidence=experimental
repeat_region	3095..3426 /rpt_family="CpG Island" /evidence=experimental
repeat_unit	complement(3663..3952) /rpt_family="Alu-Sb0 or Alu-Sb1" /evidence=experimental
repeat_unit	4912..5046 /rpt_family="Alu-J or Alu-S" /evidence=experimental
repeat_unit	5047..5333 /rpt_family="Alu-Sb0 or Alu-Sb1" /evidence=experimental
repeat_unit	5346..5628 /gene="FLN"
repeat_unit	/rpt_family="Alu-J or Alu-S" /evidence=experimental
repeat_unit	5632..5805 /rpt_family="Alu-Sq" /evidence=experimental
repeat_unit	6363..17698 /rpt_family="11Kb repeat1" /evidence=experimental
repeat_unit	6369..6633 /rpt_family="Alu-Sx" /evidence=experimental
repeat_unit	complement(7121..7409) /rpt_family="Alu-Sp" /evidence=experimental
repeat_unit	complement(7358..7432) /rpt_family="Alu-Sb0 or Alu-Sb1" /evidence=experimental
repeat_unit	complement(7426..7701) /rpt_family="Alu-J" /evidence=experimental
repeat_unit	8306..8592 /rpt_family="Alu-Sx" /evidence=experimental
repeat_unit	complement(8863..9029) /rpt_family="Alu-J or Alu-S" /evidence=experimental
repeat_unit	complement(9031..9317) /rpt_family="Alu-Sq" /evidence=experimental
repeat_unit	complement(9318..9442) /rpt_family="Alu-J or Alu-S" /evidence=experimental
repeat_unit	9486..9773 /rpt_family="Alu-J" /evidence=experimental
repeat_unit	9996..10132 /rpt_family="Alu-Sp" /evidence=experimental
repeat_unit	10159..10255 /rpt_family="Alu-Sq" /evidence=experimental
repeat_unit	10271..10367 /rpt_family="Alu-J" /evidence=experimental
repeat_unit	complement(10455..10739) /rpt_family="Alu-Sx" /evidence=experimental
repeat_unit	complement(10740..10860) /rpt_family="Alu-J or Alu-S" /evidence=experimental
repeat_unit	complement(10866..11156) /rpt_family="Alu-Sx" /evidence=experimental
repeat_unit	11223..11501 /rpt_family="Alu-J" /evidence=experimental
repeat_region	11638..13564 /rpt_family="CpG Island" /evidence=experimental
repeat_unit	complement(11884..12163) /rpt_family="Alu-Sb0 or Alu-Sb1" /evidence=experimental
repeat_unit	complement(12195..12482) /rpt_family="Alu-J or Alu-S" /evidence=experimental
repeat_unit	complement(12668..12945) /rpt_family="Alu-J or Alu-S" /evidence=experimental
repeat_unit	complement(18360..18648) /rpt_family="Alu-Sb0" /evidence=experimental
exon	complement(18984..19488) /gene="FLN"

```
/number=48
/evidence-experimental
complement(18984..45006)
/gene="FLN"
CDS
complement(join(19301..19488,19814..20017,20101..20319,
20483..20659,21361..21493,22033..22148,22336..22473,
22633..22899,23005..23127,23224..23376,23453..23656,
23748..23909,24006..24179,24367..24495,24603..24743,
24833..24935,25068..25163,25277..25524,27701..27724,
27884..28073,28649..28805,28895..29018,29434..29604,
29696..29856,29934..30096,30182..30355,30440..31037,
31758..32026,32120..32237,32429..32598,32692..32782,
32868..33028,33111..33234,34472..34615,34709..34822,
34976..35169,35271..35407,35586..35709,35798..35935,
36473..36673,36757..36919,37011..37088,37181..37299,
37846..37993,38090..38187,38291..38539,41319..41691))
/gene="FLN"
/codon_start=1
/evidence-experimental
/product="filamin"
/db_xref="GI:1203969"
/translation="MSSSRAGQSAAGAGGGVTRDAEMPATEKDLAEDAPKKI
QONTFRWNEHLKQSKRIANLQDLSGLRLIALLEVLISQKMRKHNRPTFRQM
QLENVSLVEKDSIKLVSIDSKRAVDGNLKLILGLIWLILHYSIMPMWDEED
EAKKOTPKORLLGWINKLPQITPITFSDWSGRALGALVDSAPGLCPDQSWDA
SKPTNARAAQADDLGIPQITPEEIVDPNVDHSHVMTYLSQFPKAKLPKAPLR
PKLNPKKARAYGPIEPTGMVKKRAEFTVETRSAGQEVLYVVDPAQHOREAKVLA
NNDKNRTFSVWYPEVTGTHKTVLFAQHIAKSPFVYVDKSDQDAKVTYQAQGLE
PSGNIANKTTFEITAGAGTGEVWVITDPMQKGTVEPQLEARGDSTYCSQPTM
EGVHTVHTFAGVPIRPSYTVVGCACPSACRAVGRGLQPKVRVETKADFVYTK
GAGSGELKTVKAGPEERVKQDGLDGYGVEYFYPMPGTIVITWGGNIGRSPF
EYKVGTECNVRWAGPGLGVGKSAFVVEAIGDDVGTGLGSEVSGQAKTECD
DKGDCSDYVWQAGEYAVHVLNSEDIRLSPFADIDAPQDFHDPVARGPGL
EKTAVANPKPAEFTDAKHGKAPLRVQVQDNQEGCPVEALVKDNGNYSYVPRKP
VKHTAMNSGGVYSINSPRVNAGSHPNKVKYVGPVAKTGLKAHPTFTVDCAE
ACQGVDSIGIKAPGVGPAEADIDFDIIRNDNFTVKYTPRGAGSTIIMVLAQEA
TPTSPVDRVDIIDHDNTYVKYTPVQGPVGNVYTGDDPIKSPFSVAVSPSLDS
KIKVGLGKLDVQDQDEFTVKSAGQGVKASKIVGPSAAVPCPKVPGGLGADNSV
VRLPREGPEYEVTVYDGVPPVPGPFPLEAVPTKPSKAFGPGLOGGAGSPARF
TIDTKAGTGGGLGLAVEGCEAOLECLDNGDCTCSVYPTPEGDYINILFADTHIP
GSPEKAVVPCFADSKVKCSGGLERATAGEVQGVQVDCSAGSALIELICSEAGLP
ASYVIQDHDGTHITTYIPLCPGATVTYIKGGQVPNPPSKLQVEPVDTSVQCYG
PGIEGGVFEATFEFSDARALTQTGGPHKARVANPSGNLTETVQDRDGMKVFY
YTPYEEGLSHVDVTYDGSFVPSPPFPVTEGCDSPSRVRVHGPQIGSTTNKPNFTV
ETRGAGTGLGLAVEGSEAKMSMDNKDSCSVEYIPEAGTYSLNVTYGGHOVPGS
PEKVPVHDVTDASKVKCSGGLSPGMVRANLPQSFQVDTSKAGVAPLOVKYQGPGLV
EPVDVVDNADGQTNYVPSREGPYISVLYGDEVRSPFVKVPLPHDASKVNASG
```

```
Query Match 5.7% Score 70; DB 40; Length 219447;
Best Local Similarity 68.3%; Pred. No. 2.3e-08;
Matches 97; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 666 ccaggagcacaataaacagcaggaactctctctgacttggggaggtctctac 725
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 177752 CCATGCTGAGGATGAATGAAGAGACCAAGAAAGCTTTCAGTGAGCTGGGACAGGAAAC 177693
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 726 gcaaaacaaactggctccgggacagctgggtcttcataaggagcacaacacctcaggggt 785
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 177692 GCCAAGAGCTGGCCCTCCGGGACAGCTGGGTGTTGTGGGGCCCAAGGGTGTGCGAAGAC 177633
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 786 aaagccctttgagcaggtct 807
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 177632 AAGAGCCCTTTGAGCAGGTAT 177611
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 8
AC006364 AC006364 132319 bp DNA PRI 21-DEC-1999
LOCUS Homo sapiens BAC clone GSI-146J4 from 7q31.1-31.3, complete
DEFINITION sequence.
ACCESSION AC006364
```

```
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
```

```
AC006364.3 GI:4753253
HTG.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 132319)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
2 (bases 1 to 132319)
Ryan, E., Bauer, C., Tucci, S. and Spalding, L.
The sequence of Homo sapiens BAC Clone GSI-146J4
Unpublished
3 (bases 1 to 132319)
Waterston, R.H.
Direct Submission
Submitted (11-JAN-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 132319)
Waterston, R.H.
Direct Submission
Submitted (05-MAY-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 132319)
Waterston, R.
Direct Submission
Submitted (11-SEP-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
6 (bases 1 to 132319)
Waterston, R.
Direct Submission
Submitted (21-DEC-1999) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On May 5, 1999 this sequence version replaced gi:4337266.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
----- Summary Statistics
-----
Center project name: H_GSI146J04
```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:

all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRB/CHR7>, send mail to: egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

This clone is from the first BAC library from Genome Systems, Inc. (<http://www.genomesystems.com>).


```
VERSION I66494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dorner, F., Schellinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES
    Location/Qualifiers
        1..7218
            /organism="unknown"
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

Query Match      3.8%; Score 46.6; DB 5; Length 7218;
Best Local Similarity 6.0%; Pred. No. 0.048;
Matches 25; Conservative 213; Mismatches 177; Indels 0; Gaps 0;

QY 480 atgatacagtcgtctgtgaaacaaatgtggcgagagcctaaacatcgccctgggtgaat 539
    ||| | | : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1447 AAGAATTGGTACRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1388
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 540 gaaacacggagcgtgctggacagacagcattgacatgactctgagatgttatg 599
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1387 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1328
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 600 caccatagtgaaattccttaagaaattccgggggggtgcactgtgctggcctctac 659
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1327 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1268
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 660 gacgatccaggacacaaatgaacatgaagcaggaactctctgactctggggagt 719
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1267 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1208
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 720 tctacgcaaaacactggcttcggagacgtgggtcttcctataggagcgaagacctc 779
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1207 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1148
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 780 aggggtaaaagcccttgacagctcttaagaaacagccagcacaacaaatacag 839
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1147 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1088
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 840 ggatggccagcgtcgtgagatgaggtcgatgcccgcgaagccattttagg 894
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1087 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1033
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 10
AC016137/c
LOCUS
DEFINITION Homo sapiens clone RP11-90809, *** SEQUENCING IN PROGRESS ***, 30
unordered pieces.
ACCESSION AC016137
VERSION AC016137.2 GI:7007688
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 35052)
AUTHORS Muzny, D.M., Adams, C., Bailey, M., Barbara, J., Blankenburg, K.,
Bodota, B., Bouck, J., Bowie, S., Brooks, A., Buhay, C., Bunac, C.,
Burkett, C., Burrows, J., Carter, M., Chacko, J., Chen, Z., Cox, C.,
David, R., Delgado, O., Deshazo, D., Ding, Y., Domah-Rashid, N.,
Dugan-Rocha, S., Durbin, K.J., Fernandez, C., Ferraguto, D.,
Forcum-Tansey, J., Frantz, P., Ganes, R., Gorrell, J.H., Gorrell, L.L.,
Guevara, W., Harris, K., Hernandez, J., Hodgson, A., Hoques, M.,
Holloway, C., Hosak, H., Jackson, L.E., Jackson, L., Jia, Y., Jones, M.,
Kelly, S., Kondejewski, N., Kong, Y., Kovar, C., Leal, B., Li, Z.,
Lichtarge, O., Liu, J., Liu, W., Logan, O., Lozano, R.J., Lu, J.,
Lucier, R., Martin, R., Martinez, C., McLeod, M.P., Mei, G., Morgan, M.,
```

```

TITLE Direct Submission
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 35052)
AUTHORS Worley, K.C.
JOURNAL Direct Submission
COMMENT Submitted (23-NOV-1999) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Feb 19, 2000 this sequence version replaced gi:6466486.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
-----
Center project name: HMKR
Center clone name: RP11-90809
----- Summary Statistics
Sequencing vector: M13; L08821
Chemistry: Dye-terminator Big Dye; 99% of reads
Assembly program: Phrap; version 0.980611
Consensus quality: 12430 bases at least Q40
Consensus quality: 19864 bases at least Q30
Consensus quality: 23509 bases at least Q20
Estimated insert size: 35052; agarose-fp estimation
Estimated insert size: 24454; sum-of-contigs estimation
Quality coverage: 0.3x in Q20 bases; agarose-fp estimation
Quality coverage: 0.5x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 784: contig of 784 bp in length
* 785 804: gap of unknown length
* 805 1598: contig of 794 bp in length
* 1599 1618: gap of unknown length
* 1619 2424: contig of 806 bp in length
* 2425 2444: gap of unknown length
* 2445 3277: contig of 833 bp in length
* 3278 3297: gap of unknown length
* 3298 4099: contig of 802 bp in length
* 4100 4119: gap of unknown length
* 4120 4917: contig of 798 bp in length
* 4918 4938: gap of unknown length
* 4939 5606: contig of 668 bp in length
* 5607 6721: contig of 1096 bp in length
* 6722 6741: gap of unknown length
* 6742 7433: contig of 692 bp in length
* 7434 7453: gap of unknown length
* 7454 8674: contig of 1221 bp in length
* 8675 8694: gap of unknown length
* 8695 9946: contig of 1252 bp in length
* 9947 9966: gap of unknown length
* 9967 11280: contig of 1314 bp in length
* 11281 11300: gap of unknown length
* 11301 12103: contig of 803 bp in length
* 12104 12123: gap of unknown length
* 12124 13186: contig of 1063 bp in length
* 13187 13208: gap of unknown length
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 784: contig of 784 bp in length
* 785 804: gap of unknown length
* 805 1598: contig of 794 bp in length
* 1599 1618: gap of unknown length
* 1619 2424: contig of 806 bp in length
* 2425 2444: gap of unknown length
* 2445 3277: contig of 833 bp in length
* 3278 3297: gap of unknown length
* 3298 4099: contig of 802 bp in length
* 4100 4119: gap of unknown length
* 4120 4917: contig of 798 bp in length
* 4918 4938: gap of unknown length
* 4939 5606: contig of 668 bp in length
* 5607 6721: contig of 1096 bp in length
* 6722 6741: gap of unknown length
* 6742 7433: contig of 692 bp in length
* 7434 7453: gap of unknown length
* 7454 8674: contig of 1221 bp in length
* 8675 8694: gap of unknown length
* 8695 9946: contig of 1252 bp in length
* 9947 9966: gap of unknown length
* 9967 11280: contig of 1314 bp in length
* 11281 11300: gap of unknown length
* 11301 12103: contig of 803 bp in length
* 12104 12123: gap of unknown length
* 12124 13186: contig of 1063 bp in length
* 13187 13208: gap of unknown length
-----
```

```
* 13207 14508: contig of 1302 bp in length
* 14509 14528: gap of unknown length
* 14529 15393: contig of 865 bp in length
* 15394 15413: gap of unknown length
* 15414 16603: contig of 1190 bp in length
* 16604 16623: gap of unknown length
* 16624 17853: contig of 1230 bp in length
* 17854 17873: gap of unknown length
* 17874 19035: contig of 1162 bp in length
* 19036 19055: gap of unknown length
* 19056 19920: contig of 865 bp in length
* 19921 19940: gap of unknown length
* 19941 20420: contig of 480 bp in length
* 20421 20440: gap of unknown length
* 20441 21251: contig of 811 bp in length
* 21252 21271: gap of unknown length
* 21272 22415: contig of 1144 bp in length
* 22416 22436: gap of unknown length
* 22437 23901: contig of 1466 bp in length
* 23902 23921: gap of unknown length
* 23922 25452: contig of 1531 bp in length
* 25453 25472: gap of unknown length
* 25473 26950: contig of 1478 bp in length
* 26951 26970: gap of unknown length
* 26971 29347: contig of 2376 bp in length
* 29348 29366: gap of unknown length
* 29367 31505: contig of 2139 bp in length
* 31506 31525: gap of unknown length
* 31526 32876: contig of 1351 bp in length
* 32877 32896: gap of unknown length
* 32897 35052: contig of 2156 bp in length.
FEATURES             Location/Qualifiers
    source            1..35052
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="RP11-90809"
BASE COUNT      8378 a  9047 c  7874 g  9070 t   683 others
ORIGIN
Query Match          3.5%; Score 43.4; DB 57; Length 35052;
Best Local Similarity 79.0%; Pred. No. 0.34;
Matches 64; Conservative 0; Mismatches 16; Indels 1; Gaps 1;
Qy 461 tatgtcttgaagaccgatcatcgtctgtgaaacaaatgtggcagagcct 520
Db 16366 TTTCTTTCATGCTCTCTGTTTCATGATCTCTGTAACATGTGGCAGAGCCT 16307
Qy 521 aaacatgcctcgtgtaatgg 541
Db 16306 TTT-ATCGCCCTGGTGAATGG 16287

RESULT 11
AF090946 AF090946 1302 bp mRNA PRI 12-JAN-2000
LOCUS     Homo sapiens clone H00688.
DEFINITION Homo sapiens clone H00688.
ACCESSION AF090946
VERSION   AF090946.1 GI:6690254
KEYWORDS  FLI_CDNA.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1302)
AUTHORS   Yu,Y., Zhang,C., Luo,L., Ouyang,S., Zhang,S., Li,W., Wu,J.,
           Zhou,S., Liu,M. and He,F.
TITLE     Functional prediction of the coding sequences of 50 new genes
           deduced by analysis of cDNA clones from human fetal liver
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 1302)
AUTHORS   Yu,Y., Zhang,C., Luo,L., Ouyang,S., Zhang,S., Li,W., Wu,J.,
           Zhou,S., Liu,M. and He,F.
```

```
TITLE     Direct Submission
JOURNAL   Submitted (10-SEP-1998) Dept. of Experimental Hematology, Beijing
           Institute of Radiation Medicine, 27 Taiping RD, Beijing 100850,
           P.R.China
FEATURES   Location/Qualifiers
    source            1..1302
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /clone="H00688"
                        /tissue_type="liver"
                        /dev_stage="fetus"
BASE COUNT      324 a  217 c  179 g  415 t  167 others
ORIGIN
Query Match          3.5%; Score 43; DB 51; Length 1302;
Best Local Similarity 12.0%; Pred. No. 0.45;
Matches 28; Conservative 115; Mismatches 90; Indels 0; Gaps 0;
Qy 538 atgaacacacggagctgtgtggacagaagcatttgacatgtactctgagatgta 597
Db 545 ACKSMWTTTMYRRAMCYCYCYTTWAAAAYTYYSKWWTYMWYTYMRSCMWANKR 604
Qy 598 tgcacctagtgaattctctaaagaaattccgggggtgcactggtggtggcctct 657
Db 605 AWTCYTTTMAAAWYKKKKRAWYTTTYKSMRSCYKRRGRSCYTWRYCCTTYC 664
Qy 658 acgacgatccaggaccacaaatgaacgatgaagacgagaaactctctctgactgggga 717
Db 665 YKRWKGKCYCMWKMMSMVMYTYMYKGGCYKSMRKKRAAAARKWYTYWRKTGRRKK 724
Qy 718 gttctacgcacaaactgggttcgcgacagctgggtcttcacagagcc 770
Db 725 SYKSAAMCCMRSCYTAWKRWYTTTAKGSAARSMYTTTKKKIMWKYYSI 777

RESULT 12
HSJ963E22/c
LOCUS     HSJ963E22 173823 bp DNA HTG 03-FEB-2000
DEFINITION Homo sapiens chromosome 20 clone RP5-963E22, *** SEQUENCING IN
           PROGRESS ***, 16 unordered pieces.
ACCESSION AL096828
VERSION   AL096828.18 GI:6982696
KEYWORDS  HTG; HTGS_PHASE1.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
           Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 173823)
AUTHORS   Lloyd,D.
JOURNAL   Direct Submission
           Submitted (03-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,
           CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
COMMENT   requests: clonerequest@sanger.ac.uk
           On Feb 16, 2000 this sequence version replaced gi:6911686.
           IMPORTANT: this sequence is unfinished and does not necessarily
           represent the correct sequence. Work on the sequence is in
           progress and the release of this data is based on the understanding
           that the sequence may change as work continues. The sequence may
           be contaminated with foreign sequence from E.coli, yeast, vector,
           phage etc. Order of segments is not known; 800 n's separate
           segments. Contig_ID: 00572 Length: 1314bp
           Contig_ID: 01063 Length: 9948bp
           Contig_ID: 01189 Length: 50802bp
           Contig_ID: 01224 Length: 1175bp
           Contig_ID: 02145 Length: 58359bp
           Contig_ID: 02238 Length: 1505bp
           Contig_ID: 02506 Length: 13105bp
           Contig_ID: 02914 Length: 1832bp
           Contig_ID: 04122 Length: 4757bp
           Contig_ID: 04133 Length: 1039bp
           Contig_ID: 04135 Length: 1189bp
           Contig_ID: 04137 Length: 1203bp
```

Contig_ID: 04138 Length: 1158bp
 Contig_ID: 04139 Length: 1427bp
 Contig_ID: 04160 Length: 9238bp
 Contig_ID: 04214 Length: 3772bp
 * NOTE: This is a 'working draft', sequence. It currently
 * consists of 16 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 1314: contig of 1314 bp in length
 * 1315 2114: gap of 800 bp
 * 2115 12062: contig of 9948 bp in length
 * 12063 12862: gap of 800 bp
 * 12863 63664: contig of 50802 bp in length
 * 63665 64464: gap of 800 bp
 * 64465 65639: contig of 1175 bp in length
 * 65640 66439: gap of 800 bp
 * 66440 124798: contig of 58359 bp in length
 * 124799 125598: gap of 800 bp
 * 125599 127103: contig of 1505 bp in length
 * 127104 127903: gap of 800 bp
 * 127904 141008: contig of 13105 bp in length
 * 141009 141808: gap of 800 bp
 * 141809 143640: contig of 1832 bp in length
 * 143641 144440: gap of 800 bp
 * 144441 149197: contig of 4757 bp in length
 * 149198 149997: gap of 800 bp
 * 149998 151036: contig of 1039 bp in length
 * 151037 151836: gap of 800 bp
 * 151837 153025: contig of 1189 bp in length
 * 153026 153825: gap of 800 bp
 * 153826 155028: contig of 1203 bp in length
 * 155029 155828: gap of 800 bp
 * 155829 156986: contig of 1158 bp in length
 * 156987 157786: gap of 800 bp
 * 157787 159213: contig of 1427 bp in length
 * 159214 160013: gap of 800 bp
 * 160014 169251: contig of 9238 bp in length
 * 169252 170051: gap of 800 bp
 * 170052 173822: contig of 3772 bp in length.

FEATURES

source
 1. .173823
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="20"
 /clone="RP5-963E22"
 /clone.lib="RPCI-5"

BASE COUNT 37499 a 44243 c 44350 g 35714 t 12017 others
 ORIGIN

Query Match 3.5%; Score 43; DB 33; Length 173823;
 Best Local Similarity 53.0%; Pred. No. 0.43;
 Matches 114; Conservative 0; Mismatches 100; Indels 1; Gaps 1;

QY 889 tttaggtgctgtgctcttctcagccaggccctgaagaagctcctgctgaacttag 948
 Db 71749 TCTGTGGCATCATGGCTGTGGCTGATGCTGAGTTCCAGAAAGAGCGCGGAATGACCTGC 71690
 QY 949 gagtcagagccgagggctgagggagagagagagaggggtgctgctggaaggtgctg 1008
 Db 71689 GTGCTCAGCCCTGCA-GGGATGAGGATGGCAAGGAGGGGGGAGGGGGCGGGTG 71631
 QY 1009 caggctctgacgctgtgtcgccctctcctcctcctcctcctcctcctcctcctc 1068
 Db 71630 CTGGCCGCTGCTCCCTCTGAACACATCCCGAGCCCGGGCTGGGCCCCACCGCTGCCTC 71571
 QY 1069 atctaccgggaagaccagctcagaggggtccttc 1103
 Db 71570 AGCTTCCCTGGAGAGCCTCTTCTTGATCTCATTC 71536

RESULT 13

HSJ697K14/c
 LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

HSJ697K14 123645 bp DNA HTG 11-FEB-2000
 Homo sapiens chromosome 20 clone RP4-697K14, *** SEQUENCING IN
 PROGRESS ***, 8 unordered pieces.
 ALL121829
 AL121829.8 GI:6983416
 HTG; HTGS_PHASE1; HTGS_DRAFT.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 123645)
 Sehra,H.
 Direct Submission
 Submitted (10-FEB-2000) Sanger Centre, Hinxton, Cambridgeshire,
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
 requests: clonerequest@sanger.ac.uk
 On Feb 16, 2000 this sequence version replaced gi:6977930.
 IMPORTANT: This sequence is unfinished and does not necessarily
 represent the correct sequence. Work on the sequence is in
 progress and the release of this data is based on the understanding
 that the sequence may change as work continues. The sequence may
 be contaminated with foreign sequence from E.coli, yeast, vector,
 phage etc. Order of segments is not known; 800 n's separate
 segments. Contig_ID: 00059 Length: 1315bp
 Contig_ID: 00260 Length: 1123bp
 Contig_ID: 01376 Length: 1552bp
 Contig_ID: 01477 Length: 1202bp
 Contig_ID: 02762 Length: 1123bp
 Contig_ID: 03003 Length: 48632bp
 Contig_ID: 03013 Length: 15064bp
 Contig_ID: 03016 Length: 48034bp.

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 8 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.

1 1315: contig of 1315 bp in length
 * 1316 2115: gap of 800 bp
 * 2116 3238: contig of 1123 bp in length
 * 3239 4038: gap of 800 bp
 * 4039 5590: contig of 1552 bp in length
 * 5591 6390: gap of 800 bp
 * 6391 7592: contig of 1202 bp in length
 * 7593 8392: gap of 800 bp
 * 8393 9515: contig of 1123 bp in length
 * 9516 10315: gap of 800 bp
 * 10316 58947: contig of 48632 bp in length
 * 58948 59747: gap of 800 bp
 * 59748 74811: contig of 15064 bp in length
 * 74812 75611: gap of 800 bp
 * 75612 123645: contig of 48034 bp in length.

FEATURES

source

1. .123645
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="20"
 /clone="RP4-697K14"
 /clone.lib="RPCI-4"

BASE COUNT 24500 a 35331 c 34990 g 23221 t 5603 others
 ORIGIN

Query Match 3.4%; Score 41.4; DB 33; Length 123645;
 Best Local Similarity 47.5%; Pred. No. 1.2;
 Matches 123; Conservative 0; Mismatches 136; Indels 0; Gaps 0;

